

# Penerapan Metode *Collaborative Filtering* Menggunakan *Rating* Implisit pada Sistem Rekomendasi Pemilihan Film di Rental VCD

Lisniati Dzumiroh  
Jurusan Informatika  
Universitas Sebelas Maret  
Jl. Ir. Sutami 36A Surakarta  
lisniati.dz@gmail.com

Ristu Saptono  
Jurusan Informatika  
Universitas Sebelas Maret  
Jl. Ir. Sutami 36A Surakarta  
r\_saptono@uns.ac.id

## ABSTRAK

Untuk membangkitkan rekomendasi film yang bersifat personal dan dapat sedikit di luar dugaan bagi member di suatu rental VCD, diterapkan metode *collaborative filtering recommendation*. *Collaborative filtering* memungkinkan munculnya item yang memiliki karakteristik sama sekali berbeda dari item-item yang pernah dipilih sebelumnya namun ternyata menarik bagi user bersangkutan, karena rekomendasi didasarkan pada preferensi user-user lain juga. *Feedback* ditangkap secara implisit berupa data biner dengan hanya didasarkan pada perilaku seorang member apakah dia menyewa ('1') ataukah belum menyewa ('0') suatu judul film tertentu. Metode *collaborative filtering* yang digunakan adalah *user-based collaborative filtering*, *item-based collaborative filtering*, dan *item-based collaborative filtering* yang dikombinasikan dengan fitur konten. Hasil dari pengujian ketiga metode menunjukkan bahwa pada penggunaan *user-based collaborative filtering* terjadi kesalahan prediksi rata-rata sebanyak 58,8%; pada *item-based collaborative filtering* terjadi kesalahan prediksi rata-rata sebanyak 24,9%; sedangkan pada *item-based collaborative filtering* yang dikombinasikan dengan fitur konten terjadi kesalahan prediksi rata-rata sebanyak 24,4%. Pengkombinasian *collaborative filtering* dengan fitur konten mengakibatkan hasil rekomendasi yang muncul tidak lagi memiliki karakteristik rekomendasi hasil *collaborative filtering*.

## Kata Kunci

*item-based collaborative filtering*, kombinasi fitur konten, *rating* implisit, *user-based collaborative filtering*

## 1. PENDAHULUAN

Seringkali ketika seseorang memasuki suatu rental VCD, ia mengalami kebingungan disebabkan oleh begitu banyaknya pilihan film yang tersedia. Hal seperti ini terjadi khususnya bagi mereka yang sebelumnya tidak memiliki cukup informasi seperti dari membaca *review-review* film, atau bagi mereka yang memang belum memiliki tujuan pasti akan menyewa suatu judul film tertentu. *Member-member* seperti ini tentu memerlukan bentuk rekomendasi dari *member-member* lain. Tatanan rak khusus untuk film-film yang direkomendasikan (*recommended*) pada dasarnya memuat film-film yang memang terkenal sehingga tanpa perlu direkomendasikan pun sebenarnya banyak *member-rental* VCD tersebut yang akan menyewanya. Terkadang seseorang menginginkan rekomendasi yang sedikit di luar dugaan. Kemungkinan film yang sama sekali tidak terpikirkan olehnya namun ternyata menarik untuk dilihat dan sesuai dengan selera.

Solusi dari permasalahan di atas adalah disediakannya suatu sistem rekomendasi di dalam rental VCD tersebut yang mampu menyediakan rekomendasi yang bersifat personal.

Terdapat dua pendekatan utama dalam proses rekomendasi, yaitu *content based filtering* dan *collaborative filtering*. *Content based filtering* memberikan rekomendasi item yang kemungkinan disukai oleh user tertentu berdasarkan karakteristik dari item yang sebelumnya telah mendapat penilaian positif dari user yang bersangkutan. Akibatnya, metode ini hanya dapat memberikan rekomendasi berupa *item-item* yang karakteristiknya mirip dengan *item-item* yang pernah dipilih sebelumnya. *Item* yang memiliki karakteristik sama sekali berbeda dari *item-item* yang pernah dipilih sebelumnya namun mungkin sebenarnya menarik bagi user yang bersangkutan, tidak akan pernah direkomendasikan. *Collaborative filtering* memungkinkan hal ini terjadi karena rekomendasi berdasarkan preferensi *user-user* lain juga, yaitu dengan asumsi bahwa *user-user* dengan selera yang hampir sama akan memberi penilaian terhadap *item-item* dengan cara yang hampir sama pula [1]. Dalam pembuatan sistem rekomendasi pemilihan film di rental VCD ini, dipilih menggunakan metode *collaborative filtering*.

*Feedback* dari user dalam bentuk *rating* merupakan bagian yang tidak dapat dipisahkan dari sistem rekomendasi *collaborative filtering*. *Explicit feedback* memungkinkan untuk didapatkannya deskripsi yang akurat mengenai preferensi seorang user. Namun, masalah yang dihadapi oleh cara pengumpulan *feedback* secara eksplisit adalah diperlukannya *effort* tambahan dari sisi user yang menjadikan pengumpulan *explicit feedback* menjadi relatif sukar sebab tidak semua user dengan senang hati melakukan aksi pemberian *feedback* ini. Pada kasus rental VCD, pemberian *feedback* secara eksplisit sukar diimplementasikan karena kegiatan normal di rental VCD hanya sewa dan kembali, tanpa user harus repot-repot memberikan penilaiannya terhadap film yang baru saja ia sewa. Hal ini berakibat pengumpulan *rating* dalam rentang numerik seperti pada kebanyakan sistem rekomendasi tidak mungkin untuk dilakukan. Maka, *rating* yang ditangkap secara implisit berupa data biner dengan hanya didasarkan pada perilaku seorang member apakah dia menyewa ('1') ataukah belum menyewa ('0') suatu judul film tertentu.

Pada penelitian ini, diimplementasikan metode *collaborative filtering* menggunakan *rating* implisit biner untuk menyediakan rekomendasi personal bagi *member-member* di rental VCD berupa film-film yang mungkin menarik sehingga akhirnya dapat membantu dalam proses pemilihan film-film yang ingin disewa.

## 2. SISTEM PEREKOMENDASI

Ide awal sistem rekomendasi muncul sejak berkembangnya Internet yang menimbulkan permasalahan baru yang disebut dengan *information overload*. Pengguna seringkali dihadapkan dengan situasi dimana mereka mempunyai terlalu banyak opsi

untuk dipilih, sehingga mereka membutuhkan bantuan dalam menentukan preferensi dari begitu banyaknya pilihan yang ada [2].

Sistem rekomendasi (*recommender system*) merupakan suatu agen perangkat lunak yang mempelajari minat dan preferensi seorang pengguna individu terhadap produk-produk, kemudian menyediakan rekomendasi yang sesuai dengan kebutuhan pengguna bersangkutan [3]. Sistem rekomendasi mempunyai karakteristik sebagai salah satu jenis dari *customer decision support system* (DSS) dikarenakan sistem rekomendasi: (a) merupakan sebuah sistem informasi, (b) digunakan dalam pengambilan keputusan, dan (c) digunakan untuk mendukung, bukan menggantikan, manusia dalam pengambilan keputusannya. Sama seperti jenis DSS yang lain, pada sistem rekomendasi, *customer* menyediakan masukan (berupa karakteristik produk yang diinginkan atau berupa *rating* terhadap produk) yang kemudian digunakan oleh sistem untuk membangkitkan rekomendasi bagi *customer* yang bersangkutan. Meskipun begitu, sistem rekomendasi berbeda dengan DSS dalam hal penggunaannya. Pengguna DSS biasanya adalah *general manager* atau *analyst* yang memanfaatkan sistem untuk membantu pekerjaan-pekerjaan seperti perencanaan pemasaran, perencanaan logistik, atau perencanaan keuangan, sementara pengguna sistem rekomendasi adalah *customer* yang menghadapi masalah yang disebut sebagai *preferential choice problem*.

**2.1 Collaborative Filtering Recommendation**

*Collaborative filtering* (CF) merupakan proses penyaringan atau pengevaluasian *item* dengan menggunakan opini dari orang lain. Ide utamanya adalah untuk mengeksplorasi informasi mengenai perilaku di masa lampau maupun opini dari suatu komunitas pengguna yang kemudian digunakan untuk memprediksi *item* mana yang akan disukai atau menarik bagi seorang pengguna. CF murni menggunakan matriks yang berisi *user-item rating* sebagai satu-satunya *input*, sedangkan *output* yang dihasilkan ada dua jenis: (1) prediksi (numerik) yang mengindikasikan seberapa besar tingkat kesukaan seorang pengguna terhadap sebuah *item*, dan (2) sebuah daftar berisi *item* yang direkomendasikan [4]. Istilah pengguna (*user*) dalam CF mengacu kepada mereka yang memberi penilaian terhadap *item-item* di dalam sistem, sekaligus nantinya menerima rekomendasi dari sistem.

**Tabel 1. Representasi matriks user-item rating**

User	Item1	Item2	Item3	Item4	Item5
Alice	1	1	1	1	0
User1	1	1	1	0	1
User2	1	1	0	1	0
User3	0	0	1	1	1

*Rating* dapat dikumpulkan dengan cara eksplisit, implisit, atau keduanya bersamaan. *Rating* eksplisit adalah ketika seorang pengguna secara langsung diminta untuk memberikan opini terhadap suatu *item*. *Rating* implisit mengandung arti bahwa sistem secara otomatis mendapatkan preferensi pengguna secara pasif dengan me-*monitor* aksi pengguna. Penilaian hanya didasarkan pada perilaku pengguna, misalnya ketika seorang *member* di perpustakaan memutuskan untuk meminjam suatu *item* buku maka *member* tersebut dianggap tertarik atau menyukai *item* tersebut, dan sebaliknya dianggap tidak tertarik atau tidak menyukai apabila tidak meminjamnya. Dengan cara ini, *user profile* dibentuk tanpa melibatkan *effort* tambahan dari *user*. Kekurangan dari cara ini tentu saja bahwa dugaan *feedback* bisa jadi tidak tepat [1].

**2.2 Metode dalam CF Recommendation**

Ada dua pendekatan utama dalam metode *collaborative filtering*, yaitu:

- a. *User-based collaborative filtering*

Pendekatan ini bersandar pada fakta bahwa seorang *user* mengikuti kelompok (*group*) lebih besar (yang perilaku individunya sama). Rekomendasi didasarkan pada *item-item* yang sering dibeli/disukai oleh berbagai anggota kelompok. Metode yang paling umum digunakan adalah *nearest neighbors method*. Berdasar pada *item-item* yang telah dipilih oleh tetangga terdekat seorang *user*, *item-item* yang kemungkinan akan dipilih oleh *user* tersebut di masa yang akan datang diprediksi [5]. Algoritma yang sering digunakan antara lain algoritma *Pearson correlation coefficient* (PCC) dan algoritma *vector space similarity* (VSS).

- b. *Item-based collaborative filtering*

Pendekatan ini bersandar pada relasi antar *item*, dianalisa dari informasi historis sehingga pembelian dari suatu *item* mengarahkan pembelian terhadap *item* lain (himpunan *user*). Rekomendasi didasarkan pada fakta bahwa seorang *user* cenderung memilih *item* yang mirip dengan *item-item* yang telah dipilihnya di masa lampau [6].

**2.3 Pendekatan CF untuk Data Biner**

Pendekatan *collaborative filtering*, baik *user-based* maupun *item-based*, menggunakan keseluruhan atau sampel dari basis data *user-item* untuk membangkitkan prediksi. Langkah-langkahnya secara garis besar sebagai berikut [7]:

- a. Perhitungan *similarity*

Pada *user-based collaborative filtering*, di langkah ini dihitung nilai *similarity* antar setiap pasang *user*. Ada berbagai macam algoritma untuk perhitungan *similarity*, salah satunya adalah *Pearson Correlation Coefficient* (PCC). PCC mengukur besarnya hubungan satu sama lain antara dua variabel.

Algoritma perhitungan *similarity* PCC tidak dapat digunakan pada data biner karena untuk menggunakan PCC, data yang dicari korelasinya harus berskala interval dan berdistribusi kontinu [8]. Perhitungan *similarity* yang dapat digunakan di antaranya adalah *cosine similarity* untuk data biner.

*Cosine similarity* antara *user a* dan *b* ditunjukkan sebagai berikut

$$sim(a,b) = \frac{n(A \cap B)}{\sqrt{n(A)n(B)}}$$

dengan  $n(A)$  adalah banyaknya *item* yang telah dipilih oleh *user a*,  $n(B)$  adalah banyaknya *item* yang telah dipilih oleh *user b*, dan  $n(A \cap B)$  adalah banyaknya *item* yang telah dipilih baik oleh *user a* maupun *b*.

Dari contoh matriks *user-item rating* yang ditunjukkan dalam Tabel 1, dapat dihitung *cosine similarity* antara Alice dengan User1 sebagai berikut

$$sim(Alice, User1) = \frac{3}{\sqrt{(4)(4)}} = 0,75$$

Pada *item-based collaborative filtering*, yang dihitung bukan *similarity* antar *user* melainkan *similarity* antar *item*. Sehingga jika berpedoman pada rumusan yang sama di atas,  $n(A)$  adalah banyaknya *user* yang telah memilih *item a*,  $n(B)$  adalah banyaknya *user* yang telah memilih *item b*, dan  $n(A \cap B)$  adalah banyaknya *user* yang telah memilih kedua *item a* dan *b*.

Dari contoh matriks *user-item rating* yang ditunjukkan dalam Tabel 1, dapat dihitung *cosine similarity* antara Item3 dengan Item5 sebagai berikut

$$sim(Item3, Item5) = \frac{2}{\sqrt{(3)(2)}} = 0,817$$

b. Perhitungan prediksi dan rekomendasi

Pada *user-based collaborative filtering*, di langkah ini sebuah *subset user* tetangga terdekat (*nearest neighbor*) dari seorang *user* aktif dipilih berdasar pada *similarity* mereka terhadap *user* bersangkutan. Untuk membuat rekomendasi bagi *user* aktif kita dapat menggunakan langkah-langkah berikut ini:

- (1) Menemukan *user* tetangga terdekat dengan nilai *similarity* yang paling besar.
- (2) Menghitung nilai prediksi terhadap *item-item* yang pernah dipilih oleh *user* tetangga terdekat namun belum pernah dipilih oleh *user* aktif, dengan rumus sebagai berikut

$$pred(ap) = r_u * sim(a, u)$$

dengan *a* adalah *user* aktif, *p* adalah *item* yang dihitung prediksinya, dan *r<sub>u</sub>* bernilai '1', yang menunjukkan bahwa *user* tetangga terdekat, *u*, sudah pernah memilih *item*.

Dalam contoh, prediksi bagi Alice untuk Item5 berdasar pada *user* tetangga terdekat, User1, adalah sebagai berikut

$$pred(Alice, Item5) = 1 * 0,75 = 0,75$$

Pada *item-based collaborative filtering*, di langkah ini sebuah *subset item* tetangga terdekat (*nearest neighbor*) dipilih berdasar pada *similarity* mereka terhadap *item-item* yang pernah dipilih oleh *user* aktif. Langkah yang hampir serupa dilakukan untuk membuat rekomendasi bagi *user* aktif yaitu sebagai berikut:

- (1) Menemukan *item* tetangga terdekat dengan nilai *similarity* yang paling besar.
- (2) Menghitung nilai prediksi terhadap *item-item* tetangga terdekat yang belum pernah dipilih oleh *user* aktif berdasar pada *similarity*-nya terhadap *item* yang sudah pernah dipilih oleh *user* aktif, dengan rumus sebagai berikut

$$pred(a, p) = r_a * sim(i, p)$$

dengan *a* adalah *user* aktif, *p* adalah *item* yang dihitung prediksinya, dan *r<sub>a</sub>* bernilai '1', yang menunjukkan bahwa *user* aktif sudah pernah memilih *item*.

Dalam contoh, prediksi bagi Alice untuk Item5 berdasar pada *item* tetangga terdekat, Item3, adalah sebagai berikut

$$pred(Alice, Item5) = 1 * 0,817 = 0,817$$

c. Top-N Recommendation

Dari perhitungan prediksi dan rekomendasi, didapatkan kandidat-kandidat *item* untuk direkomendasikan, yang kemudian diseleksi menjadi hanya *N* item dengan nilai prediksi yang paling tinggi (*top-N recommendation*), yang diasumsikan mempunyai nilai peluang paling besar untuk dipilih oleh *user* aktif. *N* item tersebut kemudian ditampilkan sebagai daftar *item* yang direkomendasikan untuk *user* aktif.

2.4 Feature Combination Hybrid

*Feature combination hybrid* merupakan bentuk rekomendasi yang memanfaatkan dua atau lebih masukan data dari jenis yang berbeda. Basu[9] mengajukan sebuah *feature combination hybrid* yang menggabungkan fitur-fitur *collaborative* dengan

fitur-fitur konten dari katalog *item*. Tabel 2 menggambarkan penerapan penambahan fitur konten dalam *domain* film.

Tabel 2. Fitur konten

Item	Genre
Item1	Drama/Romance
Item2	Horror/Mystery
Item3	Horror/Mystery
Item4	Horror/Mystery
Item5	Comedy

Tabel 1 menampilkan fitur *collaborative* berupa *user* beserta *rating*-nya terhadap katalog film. *Rating* dibentuk secara implisit melalui observasi terhadap *user feedback*, misalnya penyewaan. Informasi produk dibatasi pada *genre item*. Pada pendekatan *collaborative* murni, tanpa memandang fitur produk sama sekali, Alice dan User1 dengan mudah dianggap sebagai "tetangga dekat". Akan tetapi, pendekatan *feature combination* mengidentifikasi bahwa Alice, tidak seperti User1, belum pernah menyewa film ber-genre *Comedy*. Informasi ini menyebabkan Alice dan User1 berada dalam "kelompok *genre*" yang berbeda.

3. METODOLOGI PENELITIAN

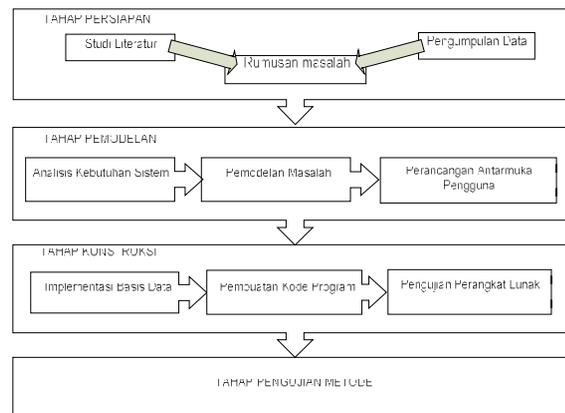
Objek penelitian dalam penulisan tugas akhir ini adalah studi kasus pada sistem rekomendasi untuk pemilihan film di rental VCD dengan data yang digunakan dalam penelitian diambil dari data transaksi dan katalog VCD di Odiva Video Rental cabang Fajar Indah Solo. Langkah-langkah yang dilakukan digambarkan pada Gambar 1.

3.1 Data Set

Sampel data diambil dari data transaksi penyewaan di Odiva Video Rental cabang Fajar Indah Solo selama bulan Januari, Februari, dan Maret 2012 sebanyak 1.345 transaksi yang melibatkan 295 item dan 218 user (tingkat *sparsity* 97,9%). *User* yang mendapatkan rekomendasi film dari sistem hanya *user* yang telah menyewa paling sedikit 3 judul film, dengan pertimbangan bahwa dengan setiap *user* minimal telah menyewa 3 *item*, formasi *neighborhood* yang masuk akal sudah dapat tercapai.

3.2 Implementasi CF untuk Binary Data

Metode *collaborative filtering* yang diimplementasikan adalah *user-based collaborative filtering*, *item-based collaborative filtering*, dan *item-based collaborative filtering* yang dikombinasikan dengan fitur konten.



Gambar 1. Diagram metode penelitian

User profile direpresentasikan dengan matriks *user-item rating* yang berisi data biner '0' atau '1'. 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. Basis data *member*, transaksi penyewaan, dan katalog *item* diubah ke dalam bentuk matriks *user-item* sebagai *training set* untuk membentuk *initial user profile*.

Pada *item-based collaborative filtering* yang dikombinasikan dengan fitur konten, suatu informasi baru ditambahkan sebagai fitur konten. Pada penelitian ini, informasi tambahan dibatasi hanya pada *genre item*. Pengklasifikasian *item* terhadap *genre* dibatasi hanya ke dalam lima macam *genre*, yaitu: *Action/Adventure*, *Drama/Romance*, *Horror/Mystery*, *Comedy*, dan *Kids/Animation*. *User* kemudian dikelompokkan sesuai dengan *genre item* yang sudah pernah disewa. Pada saat perhitungan prediksi dan rekomendasi dengan *item-based collaborative filtering*, *genre item* yang akan diprediksi terlebih dahulu dicocokkan dengan kelompok *genre* milik *user* aktif. Apabila tidak cocok, maka *item* tersebut tidak akan dihitung prediksinya sehingga tidak mungkin pula masuk dalam daftar *item* yang direkomendasikan untuk *user* bersangkutan.

Pengelompokan *user* berdasar hasil penggabungan antara fitur *collaborative* dari Tabel 1 dengan fitur konten dari Tabel 2 ditunjukkan pada Tabel 3.

Tabel 3. Pengelompokan *user* berdasar *genre item*

User	Genre
Alice	Drama/Romance, Horror/Mystery
User1	Drama/Romance, Horror/Mystery, Comedy
User2	Drama/Romance, Horror/Mystery
User3	Horror/Mystery, Comedy

### 3.3 Metrik Evaluasi

Kualitas sebuah sistem rekomendasi dapat diputuskan dari hasil evaluasi. Tipe metrik yang digunakan bermacam-macam, namun salah satu yang paling sering digunakan di literatur-literatur penelitian adalah *Mean Absolute Error* (MAE) [9]. MAE menghitung rata-rata perbedaan mutlak antara *rating* prediksi dengan *rating* sebenarnya, sebagai berikut

$$MAE = \frac{\sum_{i,j} |p_{i,j} - r_{i,j}|}{n}$$

dengan  $n$  adalah total banyaknya *rating* oleh semua *user*,  $p_{i,j}$  adalah *rating* prediksi untuk *user*  $i$  terhadap *item*  $j$ , dan  $r_{i,j}$  adalah *rating* sebenarnya. Semakin kecil nilai MAE, semakin bagus prediksi yang dihasilkan.

Metrik evaluasi lain digunakan pada pengukuran kualitas rekomendasi di penelitian ini, yaitu dengan menghitung probabilitas kesalahan prediksi. Nilai *threshold* yang digunakan adalah 0,2. Nilai tersebut dipilih mengingat tingginya tingkat *sparsity* dan besarnya data sampel yang digunakan.

Dengan menggunakan nilai *threshold* 0,2, hasil prediksi kontinu dikonversikan ke dalam bilangan biner 0 atau 1.

$$pred_{biner} = \begin{cases} 0 & : \text{if } pred_{kontinu} < 0,2 \\ 1 & : \text{if } pred_{kontinu} \geq 0,2 \end{cases}$$

Probabilitas kesalahan prediksi kemudian dihitung dengan membandingkan antara jumlah prediksi yang salah dengan jumlah prediksi keseluruhan.

## 4. PEMBAHASAN

### 4.1 Hasil Eksperimen

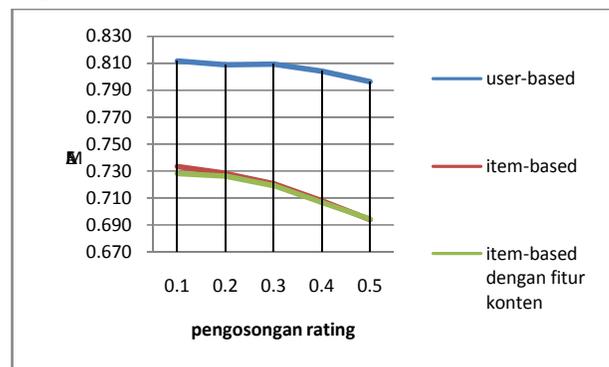
Dari data sampel yang sudah didapat, dipilih nilai *rating* untuk dikosongkan dengan menghilangkan transaksi secara acak mulai dari 10%, 20%, 30%, 40%, dan 50%. Untuk setiap data sampel yang telah dikosongkan secara acak, dihitung prediksi dengan tiga metode yang berbeda, yaitu *user-based collaborative filtering*, *item-based collaborative filtering*, dan *item-based collaborative filtering* yang dikombinasikan dengan fitur konten. Untuk masing-masing pengosongan nilai *rating* dan perhitungan nilai prediksi pada tiap metode, dilakukan percobaan sebanyak 40 kali.

Tabel 4 dan Gambar 2 menunjukkan perbandingan nilai MAE yang dihasilkan dari seluruh pengujian yang telah dilakukan.

Tabel 4. Perbandingan nilai MAE

Pengosongan rating	User-based collaborative filtering	Item-based collaborative filtering	Item-based dengan fitur konten
10%	0,812	0,733	0,728
20%	0,809	0,728	0,726
30%	0,809	0,721	0,719
40%	0,804	0,708	0,707
50%	0,796	0,694	0,694
<b>Rata-rata</b>	<b>0,806</b>	<b>0,717</b>	<b>0,715</b>

Pada level pengosongan *rating* yang semakin besar, pada umumnya akan dihasilkan nilai rata-rata MAE yang semakin besar pula dikarenakan tingkat *sparsity* semakin bertambah. Namun, pada pengujian ini nilai rata-rata MAE cenderung tidak mengalami perubahan yang berarti, atau bahkan sedikit menurun. Hal ini disebabkan karena tingkat *sparsity* yang dihasilkan pada setiap level pengosongan *rating* sebenarnya tidak banyak berubah, yaitu tetap dalam kisaran 98%. Pada pengujian yang dilakukan dengan perhitungan nilai rata-rata probabilitas kesalahan, hasil yang serupa terjadi. Nilai rata-rata probabilitas kesalahan cenderung tidak mengalami perubahan yang berarti, atau bahkan sedikit menurun.

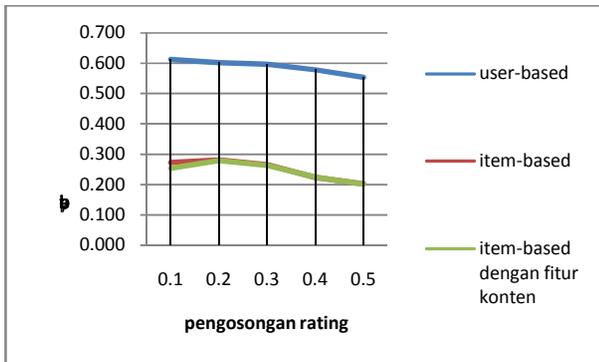


Gambar 2. Perbandingan nilai MAE

Tabel 5 dan Gambar 3 menunjukkan perbandingan nilai probabilitas kesalahan prediksi yang dihasilkan dari seluruh pengujian dengan nilai *threshold* 0,2.

Tabel 5. Perbandingan nilai probabilitas kesalahan

Pengosongan rating	User-based collaborative filtering	Item-based collaborative filtering	Item-based dengan fitur konten
10%	0,612	0,272	0,254
20%	0,602	0,281	0,278
30%	0,597	0,265	0,262
40%	0,578	0,222	0,224
50%	0,553	0,202	0,201
<b>Rata-rata</b>	<b>0,588</b>	<b>0,249</b>	<b>0,244</b>



Gambar 3. Perbandingan nilai probabilitas kesalahan

4.2 Analisis

Akurasi prediksi diukur dengan menggunakan metrik MAE dan rata-rata probabilitas kesalahan prediksi. Metrik probabilitas kesalahan prediksi digunakan karena MAE cenderung tidak sesuai digunakan untuk pengujian kualitas hasil prediksi pada sistem perekomendasi yang menggunakan data biner untuk rating-nya. MAE menggunakan rating prediksi kontinu yang bernilai antara 0 hingga 1 untuk mengukur kesalahan prediksi, padahal sistem perekomendasi yang dibangun lebih berbasis pada kejadian (occurrence) biner apakah seorang user menyewa atau belum menyewa suatu item tertentu, atau dengan kata lain lebih bersifat probabilistik.

Pengujian metode user-based collaborative filtering, item-based collaborative filtering, dan item-based collaborative filtering yang dikombinasikan dengan fitur konten pada sampel sebanyak 1.345 transaksi yang dilakukan sebanyak 40 kali dengan pengosongan rating secara acak dari 10%, 20%, 30%, 40%, dan 50% sebagaimana terlihat dalam Tabel 5 diperoleh hasil bahwa penggunaan metode item-based collaborative filtering menurunkan nilai rata-rata probabilitas kesalahan prediksi sebesar 0,339 (dari 0,588 menjadi 0,249) atau sebesar 57,7% dari nilai rata-rata probabilitas kesalahan yang dihasilkan dari penggunaan metode user-based collaborative filtering. Penggunaan metode item-based collaborative filtering yang dikombinasikan dengan fitur konten menurunkan nilai rata-rata probabilitas kesalahan sebesar 0,005 (dari 0,249 menjadi 0,244) atau sebesar 1,9% dari nilai rata-rata probabilitas kesalahan yang dihasilkan dari penggunaan metode item-based collaborative filtering. Akan Namun, pengkombinasian item-based collaborative filtering dengan fitur konten berdampak pada hasil rekomendasi yang sudah tidak lagi memiliki karakteristik rekomendasi hasil collaborative filtering. Film-film yang muncul dalam daftar rekomendasi oleh metode item-based collaborative filtering yang dikombinasikan dengan fitur konten merupakan film-film

yang ber-genre sama dengan film-film yang sudah pernah disewa oleh user bersangkutan sehingga hasil rekomendasi tidak lagi “di luar dugaan”.

Karena semakin kecil nilai kesalahan berarti semakin baik akurasi prediksi yang dihasilkan, maka dapat dikatakan bahwa metode item-based collaborative filtering menghasilkan akurasi prediksi yang lebih baik dibanding dengan metode user-based collaborative filtering, sedangkan metode item-based collaborative filtering yang dikombinasikan dengan fitur konten memberikan akurasi yang hanya sedikit lebih baik dibanding dengan metode item-based collaborative filtering.

5. KESIMPULAN

Untuk menyediakan rekomendasi film personal bagi member di rental VCD dengan berdasar pada rating implisit yang didapat, telah diimplementasikan metode collaborative filtering untuk data biner. Metode collaborative filtering yang digunakan adalah user-based collaborative filtering, item-based collaborative filtering, dan item-based collaborative filtering yang dikombinasikan dengan fitur konten.

Kualitas rekomendasi yang dihasilkan oleh penggunaan metode item-based collaborative filtering lebih baik dibandingkan dengan kualitas rekomendasi yang dihasilkan oleh penggunaan metode user-based collaborative filtering. Hal tersebut ditunjukkan dengan akurasi prediksi yang diukur dari rata-rata nilai MAE dan probabilitas kesalahan yang dihasilkan oleh metode item-based collaborative filtering lebih rendah dibandingkan dengan user-based collaborative filtering, seperti terlihat dalam Gambar 2 dan Gambar 3. Metode item-based collaborative filtering yang dikombinasikan dengan fitur konten hanya sedikit sekali memperbaiki akurasi prediksi yang dihasilkan oleh metode item-based collaborative filtering, atau dapat dikatakan perbaikan kualitas rekomendasi yang dihasilkan tidak signifikan. Selain itu, pengkombinasian collaborative filtering dengan fitur konten mengakibatkan hasil rekomendasi yang muncul tidak lagi memiliki karakteristik rekomendasi hasil collaborative filtering.

6. DAFTAR PUSTAKA

- [1] Schafer, J. B., D. Frankowski, J. Herlocker, dan S. Sen. 2007. Collaborative Filtering Recommender System. The Adaptive Web (Peter Brusilovsky, Alfred Kobsa, and Wolfgang Nejdl, eds.) Lecture Notes in Computer Science Vol. 4321 pp. 291-324. Springer - Verlag Berlin Heidelberg
- [2] Montaner, M., B. Lopez, dan J.L. De La Rosa. 2003. A Taxonomy of Recommender Agents on the Internet. Artificial Intelligence Review 19 pp.285-330. Netherlands: Kluwer Academic Publisher
- [3] Xiao, B. dan Benbasat, I. 2007. E-Commerce Product Recommendation Agents: Use, Characteristics, and Impact. MIS Quarterly 31 No. 1 pp. 137-209
- [4] Jannach, D., M. Zanker, A. Felfernig, dan G. Friedrich. 2011. Recommender Systems: An Introduction. New York: Cambridge University Press
- [5] Babu, M.S.P., dan Kumar, B.R.S. 2011. An Implementation of the User-based Collaborative Filtering Algorithm. International Journal of Computer Science and Information Technologies, Vol. 2 (3), pp. 1283-1286
- [6] Deshpande, M. dan Karypis, G. 2004. Item-based Top-N Recommendation Algorithms. ACM Transactions on Information Systems 22 No. 1 pp. 143-177

- [7] Su, X.Y. dan Khoshgoftaar, T.M. 2009. A Survey of Collaborative Filtering Techniques. *Review Article Advances in Artificial Intelligence*, Vol. 2009, Article ID 421 425, doi:10.1155/2009/421425. Hindawi Publishing Corporation
- [8] Irianto, A. 2008. *Statistika Konsep Dasar dan Aplikasinya*. Jakarta: Prenada Media Group
- [9] Basu, C., H. Hirsh, dan W. Cohen. 1998. Recommendation as Classification: Using Social and Content-Based Information in Recommendation. *Proceedings of the 15th National Conference on Artificial Intelligence*, American Association for Artificial Intelligence, pp. 714-720