

## Daftar Gambar

Gambar 1 Identifikasi Group Neighborhood .....	11
Gambar 2 Precision dan Recall .....	15
Gambar 3 Gambaran Umum Sistem.....	20
Gambar 4 Proses Preprocessing .....	21
Gambar 5 Proses Prediksi .....	24
Gambar 6 Rata-rata RMSE Uji Persamaan Prediksi .....	35
Gambar 7 RMSE terhadap seluruh grup yang dievaluasi dan target pengujian....	36
Gambar 8 RMSE terhadap seluruh grup yang dievaluasi dan target pengujian dengan group neighborhood > 0.....	36
Gambar 9 nDCG terhadap seluruh grup yang dievaluasi dan target pengujian ....	40
Gambar 10 F1-measure terhadap seluruh grup yang dievaluasi dan target pengujian .....	41

## Daftar Istilah

<i>Similarity</i>	Kemiripan, dimana dalam konteks tugas akhir ini digunakan untuk menyatakan tingkat kemiripan antar pengguna berdasarkan preferensi pengguna tersebut yang diperoleh dengan menggunakan Pearson Correlation Coefficient
<i>Neighbor</i>	Pengguna lain
<i>User Neighborhood</i>	<i>Neighbor</i> yang memiliki <i>similarity</i> yang tinggi terhadap pengguna tersebut
<i>Group Neighborhood</i>	<i>User neighborhood</i> yang muncul di semua anggota grup

# BAB 1. Pendahuluan

## 1.1. Latar Belakang

*Recommender System* pada awalnya merupakan suatu sistem yang digunakan untuk membantu pengguna secara individual dalam menentukan keputusan seperti memilih suatu buku menggunakan model dari preferensi yang telah dilakukan oleh pengguna tersebut. Seiring perkembangan bidang *Recommender System*, diketahui bahwa pada kebanyakan situasi lebih baik apabila kita dapat merekomendasikan kepada sebuah kelompok pengguna daripada pengguna secara individual [1]. Beberapa contoh pemanfaatan *Group Recommender System* seperti melihat film bersama [2], memilih perjalanan bersama [3], dan memilih stasiun radio untuk bersama misalnya pada suatu tempat *fitness* [4]. Domain yang diambil pada penelitian ini adalah buku sehingga dapat merekomendasikan buku kepada suatu grup untuk dibaca bersama.

Beberapa tahun terakhir ini, penggunaan *Recommender System* pada domain buku banyak digunakan pada *book online store* seperti Amazon dan Barnes and Noble, dari riset yang dilakukan selama 52 hari pada 156 buku didapatkan estimasi peningkatan 3% pada penjualan buku apabila ditambahkan sebuah *Recommender System* pada domain buku tersebut [5]. *Recommender System* juga berdampak kepada para pembaca dimana pada riset yang dilakukan pada *Recommender System* yang digunakan pada LibraryThing menghasilkan bahwa setelah mengecek rekomendasi yang dihasilkan, sebanyak 60% partisipan menggunakan rekomendasi yang dihasilkan, 10% hanya membaca rekomendasinya saja, 15% tidak membaca rekomendasi dan 15% lainnya [6].

*Group Recommender System* untuk buku dibuat untuk memberikan rekomendasi buku untuk dibaca bersama misalnya rekomendasi yang diberikan pada suatu komunitas baca ataupun kelompok-kelompok membaca yang kecil seperti klub membaca sehingga mereka bisa saling berbagi *review* apakah mereka menyukai maupun tidak menyukai buku tersebut. Membaca bersama dalam klub saat ini dilakukan karena memiliki beberapa manfaat antara lain membentuk pertemanan dan mendapatkan pengalaman membaca dari orang lain, menstimulasi intelektual, sebagai tempat istirahat dari kesibukan sehari-hari, dan lain-lain [7].

Rekomendasi untuk sebuah grup lebih rumit daripada merekomendasikan untuk individual karena perlu adanya cara untuk menggabungkan model-model yang dibangun dari setiap individu [1]. Model individu yang dimaksud adalah model yang dihasilkan oleh *Recommender System* setiap individu. Saat ini ada dua cara utama pada *Group Recommender System* yang sedang dikembangkan. Teknik pertama yaitu menggabungkan profil setiap individu pada masing-masing kelompok menjadi sebuah representasi dari kelompok tersebut dan yang kedua adalah menggabungkan hasil rekomendasi dari setiap individu dalam kelompok tersebut menjadi satu kesatuan rekomendasi [1].

Pendekatan yang akan dilakukan menggunakan salah satu teknik yang paling sering digunakan yaitu *Collaborative Filtering* yang menggunakan informasi dan preferensi dari pengguna lain untuk melakukan rekomendasi. Teknik *Collaborative Filtering* yang digunakan adalah *memory based* yang diambil pada dengan cara mencari preferensi pengguna yang belum diketahui dengan mencari kesamaan dengan preferensi pengguna lain [8].

Pengembangan dari teknik agregasi yang digunakan saat ini salah satunya adalah adanya penambahan prioritas pada *Group Neighborhood*, dimana *Group Neighborhood* merupakan tetangga-tetangga yang berada di seluruh ketetanggan anggota dalam suatu grup. Hasil yang diperoleh pada suatu kasus membuktikan bahwa menggunakan *Group Neighborhood* ini dapat menurunkan tingkat kesalahan yang dihasilkan dari 0,4960 menjadi 0,4956 [8].

## 1.2. Rumusan Masalah

Berdasarkan latar belakang yang telah dijelaskan, maka didapatkan rumusan masalah sebagai berikut :

1. Bagaimana mengimplementasikan teknik agregasi dan implementasi *Group Neighborhood Selection* pada preferensi individu dalam grup untuk *Memory-based Group Recommender System* pada domain buku ?
2. Teknik agregasi mana yang paling sesuai untuk diimplementasikan pada *Memory-based Group Recommender System* untuk domain buku pada dataset buku yang diambil dari Amazon berdasarkan tingkat kesalahan dalam melakukan prediksi ?

Sedangkan batasan masalah pada penelitian ini adalah :

1. *Dataset* yang digunakan adalah dataset buku pada Amazon [9].
2. Fokus utama dalam penelitian ini adalah pada analisis teknik penggabungan individual model berdasarkan *Memory-based Group Recommender System*.
3. Pengguna memiliki setidaknya 20 *rating* pada *item dataset* [10].

### 1.3. Tujuan

Tujuan dari penelitian ini adalah :

1. Mengimplementasikan teknik agregasi dan *Group Neighborhood Selection* pada preferensi setiap individu dalam grup untuk *Memory-based Group Recommendation System* pada domain Buku.
2. Memperoleh teknik agregasi yang tepat berdasarkan analisis tingkat kesalahan dengan teknik RMSE terhadap hasil prediksi dan rekomendasi yang dihasilkan serta analisis dampak dari implementasi *Group Neighborhood Selection* pada *group recommendation*.

### 1.4. Metodologi Penyelesaian Masalah

Metode yang digunakan pada penelitian ini terdapat enam tahapan, antara lain :

1. Studi Literatur  
Penelitian dimulai dengan mempelajari permasalahan yang ada pada *Group Recommender System* dan mempelajari metode-metode dan evaluasi yang dapat digunakan untuk menyelesaikan permasalahan tersebut.
2. Analisis dan Perancangan Sistem  
Menganalisis kebutuhan dan perancangan arsitektur sistem yang digunakan dalam menyelesaikan permasalahan yang diangkat.
3. Implementasi dan Pengujian Sistem  
Mengimplementasikan sistem *Group Recommendation System* yang telah dirancang dan menguji dengan menggunakan *root-mean-square error* (RMSE) guna mengukur akurasi dari prediksi hasil

rekomendasi, serta pengujian menggunakan *normalize discounted cumulative gain* (nDCG) dan *F1-measure* guna menguji hasil rekomendasi yang diberikan.

4. Analisis Hasil Pengujian

Pada tahap ini dilakukan analisis dari hasil keluaran sistem untuk menyelesaikan permasalahan yang diangkat.

5. Pembuatan Laporan

Tahap ini merupakan tahap akhir dari penelitian dimana dilakukan penyusunan laporan, dokumentasi dan kesimpulan berdasarkan analisis hasil pengujian.

## 1.5. Sistematika Penulisan

Garis besar sistematika penulisan pada buku Tugas Akhir ini adalah sebagai berikut.

1. Bab 1 Pendahuluan

Bab ini berisi gambaran mengenai latar belakang, identifikasi dan perumusan masalah, identifikasi batasan/lingkup masalah, tujuan yang ingin dicapai pada akhir penelitian, serta metodologi penyelesaian masalah tersebut dan sistematika penulisan.

2. Bab 2 Studi Literatur

Bab ini berisi teori-teori yang digunakan dalam mendukung analisis dari sistem yang dibangun berdasarkan masalah yang dihadapi, dimana di dalamnya membahas mengenai *recommender system*, *group recommender system*, *collaborative filtering*, *group neighborhood*, strategi agregasi, *pseudo user*, *root mean square error* (RMSE), *normalize discounted cumulative gain* (nDCG), dan *F1-measure*.

3. Bab 3 Perancangan Sistem

Bab ini berisi desain rancangan yang digunakan pada saat dilakukannya analisis berdasarkan studi literatur yang telah dilakukan.

4. Bab 4 Pengujian dan Analisis

Bab ini berisi deskripsi skenario yang lebih detail dan hasil pengujian berdasarkan rancangan skenario pengujian yang telah dibuat beserta analisis dari pengujian tersebut.

5. Bab 5 Kesimpulan dan Saran

Bab ini berisi kesimpulan dari keseluruhan hasil pengerjaan TA yang mengacu pada tujuan penelitian, skenario pengujian, dan analisis hasil pengujian pada bab-bab sebelumnya.

## BAB 2. Studi Literatur

### 2.1. Recommender System

*Recommender System* merupakan sistem perangkat lunak dan teknik yang digunakan untuk memberikan saran atau rekomendasi yang akan berguna bagi pengguna [11]. Rekomendasi tersebut bertujuan untuk membantu dalam pengambilan keputusan seperti *item* apa yang akan dibeli, musik apa yang akan didengarkan, atau berita apa yang sesuai untuk dibaca. Sistem rekomendasi juga memiliki peran penting dalam beberapa penyedia layanan seperti *online store* untuk meningkatkan penjualan produk mereka, salah satu contohnya adalah Amazon menggunakan *book recommender system* untuk merekomendasikan buku yang akan dibaca [11].

Beberapa metode atau strategi yang digunakan pada *Recommender System* saat ini yaitu *Collaborative Filtering*, *Content Based*, dan *Trust Based* [12]. *Collaborative Filtering* membuat rekomendasi dengan mengkombinasikan pengetahuan dari *similar user* di dalam sistem. Kelebihannya adalah mudah dibuat dan digunakan, dapat menjelaskan hasil yang diperoleh, dapat menambahkan data baru dengan mudah, dan lebih bisa diaplikasikan dalam dunia nyata. Kekurangannya adalah bergantung pada *rating* pengguna, adanya *sparsity* atau kekurangan data, tidak memperhatikan keterhubungan sosial antar pengguna, dan memiliki masalah pada *cold start user*. *Content Based* membuat rekomendasi dengan kesamaan item terhadap item yang telah dibeli atau ditinjau oleh pengguna tersebut. Kelebihannya adalah pengguna bebas memilih item dan adanya transparansi. Kelemahannya adalah analisis konten yang terbatas dan spesialisasi yang berlebihan. *Trust Based* membuat rekomendasi dengan melihat rekomendasi dari teman atau seseorang yang diketahui dan dipercaya oleh pengguna tersebut. Kelebihannya adalah dapat mengatasi *cold start user* dan mengurangi *sparsity*. Kelemahannya adalah sulit untuk membangun *trust network* [12]. Berdasarkan metode-metode yang dijelaskan, metode yang akan digunakan adalah *Collaborative Filtering* karena digunakan untuk *Memory-based Group Recommender System* dimana *Memory-based* merupakan salah satu tipe metode

*Collaborative Filtering* yang menggunakan data preferensi atau *rating* dari pengguna untuk membuat rekomendasi.

## 2.2. Group Recommender System

*Group Recommender System* merupakan sistem rekomendasi dimana target dan data preferensi rekomendasi produk atau item berupa sekelompok pengguna yang membentuk grup-grup. Terdapat banyak situasi dimana penggunaan rekomendasi terhadap grup lebih sesuai dibandingkan rekomendasi terhadap individual seperti MUSICFX yang merekomendasikan stasiun radio untuk bersama [4], POLYLENS yang merekomendasikan film untuk bersama, rekomendasi perjalanan bersama [3], dan lain-lain. *Group Recommender System* memiliki berbagai macam strategi agregasi untuk membentuk suatu model grup dan juga berbagai macam klasifikasi untuk *group recommendation* [1].

## 2.3. Collaborative Filtering

*Collaborative Filtering* (CF) merupakan salah satu metode yang paling sering digunakan dan mudah diterapkan dalam kebanyakan studi literatur. Dasar dari metode ini merupakan pemberian rekomendasi menggunakan pencocokan kesamaan preferensi dari pengguna lain [11]. Preferensi dapat diperoleh berdasarkan *rating* yang secara eksplisit dari pengguna terhadap produk maupun berupa aktivitas interaksi pengguna di dalam sistem. Preferensi yang diperoleh ini digunakan untuk membuat prediksi preferensi pengguna yang masih belum diketahui [8].

Tabel 1 Contoh data pada Collaborative Filtering

		Buku			
		Classical Mythology	Clara Callan	Decision in Normandy	The Kitchen God's Wife
User	Sari	2	3	1	?
	Budi	4	2	?	3
	Farhad	?	3	2	3
	Bimo	4	?	2	3



Tabel 1 menggambarkan data *rating* dimana tanda tanya “?” merupakan preferensi yang masih belum diketahui dan *rating* dari 1(tidak suka) hingga 4(paling suka). Sebagai contoh, Sari memberikan *rating* pada item buku “Classical Mythology”, “Clara Callan”, dan “Decision in Normandy”. Sistem dapat memprediksi preferensi Sari yang masih belum diketahui seperti *rating* pada “The Kitchen God’s Wife” menggunakan preferensi dari pengguna lain seperti Budi, Farhad, dan Bimo dengan suatu perhitungan tertentu. *Collaborative Filtering* memiliki tiga pendekatan yaitu *Memory-based CF*, *Model-based CF*, dan *Hybrid recommender* [13].

Tabel 2 Tinjauan Pendekatan CF

Pendekatan CF	Representasi teknik	Kelebihan	Kekurangan
<i>Memory-based CF</i>	- <i>Neighbor-based CF</i> dimana merupakan item-based atau pengguna-based yang menggunakan Pearson ataupun cosine correlation - <i>top-N recommendation</i> dari <i>item-based</i> dan <i>user-based</i>	- mudah diimplementasikan - data baru mudah untuk ditambahkan secara <i>increment</i> - tidak perlu mempertimbangkan isi dari item yang direkomendasikan - dapat diskalakan dengan baik dari <i>rating</i> pengguna	- bergantung pada <i>rating</i> pengguna - performansi menurun ketika data sparse - tidak dapat merekomendasikan untuk pengguna dan item baru - memiliki skalabilitas terbatas untuk <i>dataset</i> yang besar
<i>Model-based CF</i>	- Bayesian CF - Clustering CF - CF menggunakan teknik reduksi	- lebih baik dalam mengatasi data sparse, skalabilitas, dll.	- pembangunan model yang mahal - terdapat <i>trade-off</i> antara performansi dan skalabilitas sistem

	dimensi seperti PCA	- meningkatkan performansi prediksi - dapat memberikan alasan pada rekomendasi yang diberikan	- dapat menghilangkan informasi yang berguna ketika menggunakan teknik pengurangan dimensi
<i>Hybrid recommender</i>	- Content-based CF - menggabungkan Memory-based CF dan Model-based CF	- mengatasi keterbatasan CF - meningkatkan performansi prediksi - mengatasi masalah CF seperti data sparse	- dapat meningkatkan kompleksitas dan biaya untuk pelaksanaan - memerlukan informasi eksternal yang biasanya tidak tersedia

Pendekatan yang akan digunakan dalam penelitian ini adalah *Memory-based* CF, dimana merupakan salah satu pendekatan CF yang menggunakan seluruh atau sebagian data preferensi pengguna untuk membuat prediksi dengan melakukan suatu perhitungan *similarity* terhadap pengguna lain [13]. Salah satu pengukuran yang digunakan adalah Pearson Correlation, yaitu teknik perhitungan *similarity* yang digunakan untuk mengukur keterhubungan dua variabel secara linear. Sebagai contoh misal  $u$  dan  $v$  adalah pengguna, maka perhitungan *similarity* menggunakan Pearson Correlation (PCC) menggunakan rumus sebagai berikut :

$$PCC = w_{u,v} = \frac{\sum_{i \in I} (r_{u,i} - \bar{r}_u)(r_{v,i} - \bar{r}_v)}{\sqrt{\sum_{i \in I} (r_{u,i} - \bar{r}_u)^2} \sqrt{\sum_{i \in I} (r_{v,i} - \bar{r}_v)^2}} \quad (1)$$

dimana  $i \in I$  merupakan penjumlahan setiap item yang pengguna  $u$  dan  $v$  telah berikan *rating* dan  $\bar{r}_u$  adalah rata-rata *rating* setiap item oleh pengguna  $u$  [13].

Penghitungan nilai prediksi untuk pengguna tersebut digunakan persamaan sebagai berikut.

$$prediksi = p_{a,i} = \bar{r}_a + \sigma_a \frac{\sum_{b=1}^n [(r_{b,i} - \bar{r}_b) \cdot w_{a,b}] / \sigma_b}{\sum_{b=1}^n |w_{a,b}|} \quad (2)$$

dimana  $\bar{r}_a$  adalah rata-rata *rating* setiap item oleh pengguna a,  $r_{b,i}$  adalah rating item i oleh pengguna b, dan  $\sigma_b$  adalah standar deviasi rating setiap item oleh b.

Sedangkan untuk *Group Neighborhood Selection* model, terdapat penambahan dari persamaan (2) menjadi :

$$prediksiGNS1 = p_{a,i} = \bar{r}_a + \sigma_a \frac{\sum_{b=1}^n [(\alpha(r_{b,i} - \bar{r}_b) \cdot w_{a,b})] / \sigma_b}{\sum_{b=1}^n |w_{a,b}|} \quad (3)$$

dimana  $\alpha=1$  apabila *neighbor*  $\in$  *User\_Neighborhood* dan  $\alpha=2$  apabila *neighbor*  $\in$  *Group\_Neighborhood* [8]. Penyesuaian dilakukan terhadap persamaan (3) dimana penambahan bobot  $\alpha$  hanya dikalikan pada pembilang, hal itu dapat membuat *rating* pada *group neighborhood* yang semula misalkan 4 dengan rata-rata 3 yang seharusnya bernilai 1 menjadi 2, sementara berdasarkan studi literatur yang dilakukan, penambahan bobot untuk *Collaborative Filtering* seperti yang ditunjukkan oleh Y. Ding dan X.Li pada penelitiannya yang berjudul Time weight collaborative filtering [14], menggunakan persamaan:

$$O_{ij} = \frac{\sum_{c=1}^k O_{ic} \cdot sim(I_j, I_c) \cdot f(t_{ic})}{\sum_{c=1}^k sim(I_j, I_c) \cdot f(t_{ic})} \quad (4)$$

dimana bobot pada persamaan tersebut merupakan fungsi waktu  $f(t_{ic})$ . Persamaan tersebut menunjukkan bahwa penambahan perkalian untuk bobot tidak hanya pada pembilang, namun juga pada penyebut, sehingga penyesuaian dilakukan pada persamaan (3) menjadi:

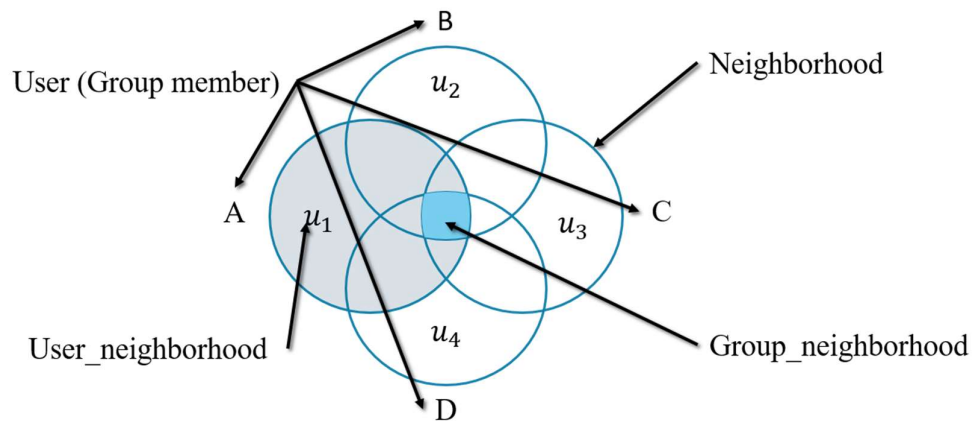
$$prediksiGNS2 = p_{a,i} = \bar{r}_a + \sigma_a \frac{\sum_{b=1}^n [(\alpha(r_{b,i} - \bar{r}_b) \cdot w_{a,b})] / \sigma_b}{\sum_{b=1}^n |\alpha \cdot (w_{a,b})|} \quad (5)$$

Persamaan yang dihasilkan memberikan bobot yang berdampak pada bertambahnya peran *group neighborhood* pada rekomendasi dilihat dari kemunculan *group neighborhood* pada proses prediksi sebanyak  $\alpha = 2$  daripada *user neighborhood* dengan  $\alpha = 1$ . Maksimum *neighborhood* yang disarankan

adalah 20 hingga 60, dalam penelitian ini digunakan maksimum *neighborhood* yang digunakan untuk prediksi sebanyak 50 *neighborhood* [8].

## 2.4. Group Neighborhood

*Group neighborhood* merupakan irisan dari ketetanggaan yang dimiliki oleh setiap pengguna dalam sebuah group. *Group neighborhood* ini nantinya akan diberikan prioritas lebih tinggi daripada *user neighborhood*, dimana *neighborhood* merupakan tetangga dari pengguna dalam suatu grup atau dengan kata lain tetangga merupakan pengguna lain yang memiliki *similarity* terdekat dengan pengguna tersebut.



Gambar 1 Identifikasi Group Neighborhood

## 2.5. Strategi Agregasi

Masalah utama *group recommendation* adalah bagaimana melakukan penyesuaian rekomendasi terhadap suatu grup pengguna berdasarkan informasi preferensi individual pengguna [1]. Beberapa *group recommendation* menggunakan teknik agregasi yang berbeda yang sesuai dengan data domain yang digunakan, misal pada MUSICFX menggunakan *average without misery* dan POLYLENS menggunakan *least misery* [1]. Sebagai contoh, misalkan ada data suatu group berisi tiga pengguna yaitu Sari, Farhad, dan Bimo serta sistem mengetahui bahwa mereka berada di dalam sistem dan mengetahui juga ketertarikan masing-masing pengguna tersebut di masing-masing item.