

Bayesian Personalized Ranking – Matrix Factorization untuk Rekomendasi Buku

Fikri Bahiransyah¹, Dade Nurjanah, Ir., M. T., Ph. D.², Rita Rismala, S. T., M. T.³

^{1,2,3}Fakultas Informatika, Universitas Telkom, Bandung

¹bahiransyah@students.telkomuniversity.ac.id, ²pembimbing1@telkomuniversity.ac.id,

²pembimbing2@telkomuniversity.ac.id

Abstrak

Pada umumnya, sistem merekomendasi sejumlah barang yang dibutuhkan oleh pengguna, terutama dalam merekomendasikan sebuah buku, namun terkadang sistem rekomendasi merekomendasikan beberapa buku yang tidak dibutuhkan oleh pengguna. Penyebab sebuah sistem rekomendasi merekomendasikan buku yang tidak sesuai kebutuhan pengguna, salah satunya dikarenakan sistem rekomendasi tersebut berfokus hanya pada rating buku dan tidak pada personal atau kepribadian pengguna. Metode Bayesian Personalized Ranking (BPR) merupakan metode yang berfokus *ranking item* yang berdasarkan pada *item* yang lebih disukai atau diketahui oleh user, dibandingkan menerapkan teknik dengan memprediksi rating. Pada tugas akhir ini, akan menerapkan metode Bayesian Personalized Ranking (BPR) dalam sistem rekomendasi buku, dataset yang akan digunakan dalam sistem menggunakan dataset Goodreads dan Bookcrossing. Dataset yang digunakan akan dimodelkan menggunakan teknik Matrix Factorization (MF). Tugas akhir ini bertujuan untuk mengetahui nilai Area Under Curve (AUC) yang dihasilkan dengan menggunakan metode BPR-MF. Dari hasil tugas akhir ini didapatkan nilai skor $AUC(u) = 0.962$ untuk dataset Goodreads dan $AUC(u) = 0.95518$ untuk dataset Bookcrossing. Dari skor yang didapat menjelaskan bahwa metode BPR-MF bekerja dengan baik dalam memprediksi buku untuk direkomendasikan kepada user.

Kata kunci : sistem rekomendasi, bayesian personalized ranking, matrix factorization, buku

Abstract

In general, the recommendation system is recommending a number of items that the user needs, especially in recommending a book, but sometimes recommendation systems is recommending some books that are not required by the user. The cause of a recommendation system recommends books that do not fit the needs of users, which is because the recommendation system focuses only on book ratings and not on personal or user personalities. The Bayesian Personalized Ranking (BPR) method is a method of ranking-based items that are based on favored items or items that known by the user, rather than applying a technique by predicting a rating. In this final project, will implementing Bayesian Personalized Ranking (BPR) method in recommendation system for recommending books, the dataset will be used in the system using Goodreads and Bookcrossing dataset. Dataset will be modeled using the Matrix Factorization (MF) technique. This final project aims to find out the value of Area Under Curve (AUC) produced by using BPR-MF method. From the results of this final task is obtained value of $AUC(u) = 0.962$ for datasets Goodreads and $AUC(u) = 0.95518$ for Bookcrossing datasets. From the scores obtained explained that the BPR-MF method works well in predicting books to be recommended to users..

Keywords: recommendation system, bayesian personalized ranking, matrix factorization, book

1. Pendahuluan

1.1. Latar Belakang

Semakin berkembangnya jaman, maka semakin banyak ragam buku yang beredar di dunia maya. Dengan adanya sistem rekomendasi situs buku *online* dapat menyediakan buku yang relevan dan dibutuhkan sesuai dengan preferensi seseorang. Hampir semua situs buku *online* sudah menggunakan sistem rekomendasi seperti *Google*, *Scribd*, dan *Openlibrary*

Pada umumnya, situs buku *online* yang telah menerapkan sistem rekomendasi namun, tidak selalu menghasilkan rekomendasi yang diharapkan oleh user. Hal tersebut disebabkan karena, sistem rekomendasi memprediksi rekomendasi produk yang memiliki rating tinggi atau berdasarkan riwayat orang yang mengakses

jenis buku tertentu pada situs *online*. Sistem rekomendasi yang tidak memperhatikan seluruh preferensi pengguna dapat mengakibatkan rekomendasi yang tidak akurat.

Pada tugas akhir ini, metode Bayesian Personalized *Ranking* dengan pemodelan data menggunakan Matrix Factorization akan diimplementasikan untuk membangun Sistem Rekomendasi Buku yang dapat mengatasi permasalahan yang terjadi pada situs buku online, dengan Collaborative Filtering sebagai basis sistem rekomendasi yang digunakan.

1.2. Topik dan Batasan

Collaborative Filtering (CF) [2] sudah digunakan diberbagai macam sistem yang memiliki tujuan yang berbeda-beda, hal ini dikarenakan CF memiliki kesuksesan lebih besar dalam menghasilkan rekomendasi yang akurat dibandingkan Content-based [2]. Saat ini metode yang terkenal dalam CF merupakan Matrix Factorization (MF) [5][4], dikarenakan MF dapat memodelkan data dari pengguna dan data barang, sehingga MF dapat mengenali karakteristik kedua jenis data tersebut.

Pada tugas akhir ini penulis menggunakan metode Bayesian Personalized *Ranking* – Matrix Factorization (BPR-MF) [1], untuk menghasilkan rekomendasi yang akurat dan relevan terhadap ketertarikan seorang pengguna. BPR-MF [1] merupakan pengembangan dari metode MF, yang berfokus pada *ranking* suatu *item* berdasarkan *feedback* pengguna untuk direkomendasikan pada pengguna lainnya. Pada BPR-MF[1] setiap user yang berinteraksi dengan item tertentu maka, item tersebut dapat menghasilkan nilai prediksi rekomendasi dibandingkan item lainnya yang jarang berinteraksi dengan user. Contohnya pada tugas akhir ini yakni, buku cerita lebih banyak dibaca oleh user dibandingkan buku pelajaran sehingga, hasil prediksi buku cerita akan menghasilkan nilai yang lebih tinggi dibandingkan buku pelajaran. Pada sistem akan diranking nilai prediksi tertinggi suatu user yang nantinya akan dipakai untuk user lainnya. Tugas akhir ini menggunakan metode BPR-MF yang memiliki akurasi yang baik [6] disebabkan oleh perhitungan kemungkinan setiap data berdasarkan dependensinya, dan BPR-MF dapat memodelkan berbagai macam keadaan dalam kehidupan nyata secara fleksibel [5]. Rumusan masalah yang didapat pada tugas akhir ini yakni :

1. Bagaimana cara mengimplementasikan metode Bayesian Personalized *Ranking* – Matrix Factorization untuk merekomendasikan buku?
2. Bagaimana pengaruh metode Bayesian Personalized *Ranking* – Matrix Factorization pada akurasi rekomendasi buku?

Kemudian terdapat batasan – batasan masalah pada tugas akhir ini yang bertujuan untuk menyesuaikan kebutuhan serta kemampuan penulis, yakni :

1. Dataset yang digunakan yaitu dataset Goodreads dan Bookcrossing, karena dataset pada goodread dan bookcrossing mudah didapatkan dan lebih mendukung metode yang berfokus pada preferensi pengguna [9] [10].
2. Pada tugas besar ini, data yang akan dimodelkan merupakan data yang bersifat implisit atau berisikan interaksi antara item dan user, berdasarkan riwayat user yang memberikan rating kepada sebuah buku sehingga, semua data akan bersifat feedback positif, sedangkan user yang tidak memiliki riwayat dengan sebuah buku akan dimodelkan menjadi feedback negatif [8].

1.3. Tujuan

Berdasarkan permasalahan yang terdapat pada latar belakang yang dijelaskan dan pada bagian sebelumnya, maka didapatkan tujuan dari tugas akhir ini sebagai berikut :

1. Membangun sistem untuk merekomendasikan buku berdasarkan hasil prediksi dari metode *Bayesian Personalized Ranking – Matrix Factorization*.
2. Mengetahui besar akurasi untuk memprediksi buku yang akan direkomendasi dengan metode *Bayesian Personalized Ranking – Matrix Factorization*.

1.4. Organisasi Tulisan

Berikut ini merupakan susunan sistematika tugas akhir ini :

1. Pendahuluan
Pada bagian ini berisikan keterangan atau penjelasan mengenai latar belakang, topik tugas akhir, batasan, tujuan dan organisasi tulisan tugas akhir ini.
2. Studi Terkait
Pada bagian ini berisikan studi literatur yang berhubungan serta mendukung tugas akhir ini.
3. Sistem yang Dibangun

Pada bagian ini berisikan penjelasan sistem yang dibangun, sekaligus cara kerja dan alur kerja sistem tersebut.

4. Evaluasi

Pada bagian ini berisikan hasil pengujian dari proses yang dilakukan pada tugas akhir ini.

5. Kesimpulan

Pada bagian ini berisikan kesimpulan dari tugas akhir ini.

2. Studi Terkait

2.1. Recommender System

Recommender System atau sistem rekomendasi merupakan kepingaran ciptaan pada sistem dengan menggunakan teknik data *mining* dan *mechine learning*[2], yang tujuannya untuk membantu pengguna dalam memutuskan suatu hal berdasarkan kebiasaan pengguna tersebut. Sehingga sistem rekomendasi menjadi suatu teknologi yang dapat menyaring informasi untuk memprediksi, ketertarikan seseorang terhadap suatu benda[2].

Sistem Rekomendasi sudah diterapkan diberbagai macam hal, khususnya pada sistem pencarian seperti google. Pada sistem pencarian google, biasanya sistem rekomendasi dapat dilihat saat pengguna akan mencari mengenai suatu hal dan google akan memberikan sejumlah pilihan yang menurut sistem, pengguna akan memilih pilihan tersebut[2][5].

Sedangkan pada penelitian [7], dilakukan pembangunan sistem rekomendasi untuk merekomendasikan buku menggunakan metode Collaborative Filtering(CF). Langkah-langkah yang dilakukan penelitian ini untuk merekomendasikan buku antara lain, mengumpulkan data dan diorganisir kedalam database. Kemudian sistem akan mempelajari pilihan yang biasa dilakukan user. Kemudian seluruh informasi yang didapatkan, sistem akan memberikan rekomendasi buku menggunakan hasil dari CF, hasil dari CF yakni rekomendasi yang diberikan berdsarkan preferensi dua user yang sama terhadap suatu *item*. Hasil yang didapatkan pada penelitian [7] yakni akurasi sebesar 77% dengan minimal sepuluh buku yang dipilih oleh user

2.2. Collaborative Filtering (CF)

CF merupakan salah satu pendekatan dari sistem rekomendasi, yang berfokus dalam menyaring suatu hal berdasarkan kesamaan antar pengguna, yang nantinya akan direkomendasikan. CF akan merekomendasikan barang atau hal lainnya berdasarkan nilai dari barang tersebut. Selain dari faktor barang, CF juga mempertimbangkan kesamaan pengguna berdasarkan kebiasaannya, sehingga dapat menghasilkan rekomendasi barang yang akurat [2].

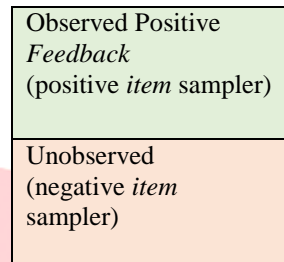
CF memiliki matriks user-*item* untuk menemukan pengguna yang memiliki ketertarikan yang sama. CF memiliki kelebihan dalam memperhitungkan data, karena data yang digunakan dalam menentukan similarity atau kesamaan antara pengguna dapat berdasarkan, kebiasaan pengguna atau berdasarkan barang itu sendiri, yang nantinya akan dimasukkan kedalam matriks user-*item* [2]. CF pun sudah diterapkan dalam beberapa hal, contohnya dalam merekomendasikan film, online shop, dan rental musik [2][4].

2.3. Bayesian Personalized Ranking (BPR)

BPR merupakan sistem rekomendasi yang merekomendasikan sejumlah barang berdasarkan peringkat barang tersebut. Contohnya penjualan buku secara online yang merekomendasikan beberapa kategori buku, berdasarkan kategori yang diminati pembeli. Selain itu, BPR mengolah data sample atau data traning untuk rekomendasi barang yang hanya bersifat data negatif atau positif. Catatan yang data yang diperhatikan hanyalah data yang bersifat positif, data yang tersisa akan diklasifikasikan oleh model sebagai data negatif. Data negatif tersebut dapat diakibatkan karena, pengguna mungkin tertarik dengan barang tersebut, atau barang tersebut tidak diketahui oleh pengguna [4].

BPR memiliki kekurangan dalam memodelkan data, sehingga BPR membutuhkan suatu metode yang dapat memodelkan memodelkan sperti *Matrix Factorization*(MF). Namun BPR memiliki kelebihan dalam melakukan sampling data, karena BPR dapat memasang sejumlah data secara optimal, serta BPR dapat mempelajari tingkat hubungan antara dua barang dan membuat urutan dari besar hubungan kedua barang tersebut [6].

BPR sudah diimplementasikan dalam sistem yang berfungsi untuk merekomendasikan pekerjaan pada seseorang. Dalam sistem tersebut, BPR memiliki akurasi yang tinggi dikarenakan sampling data pada BPR menetapkan barang yang di observasi sebagai nilai positif, sedangkan yang lain akan ditetapkan sebagai nilai negatif [6].



Gambar 1. Contoh *sampling* pada Metode BPR[6].

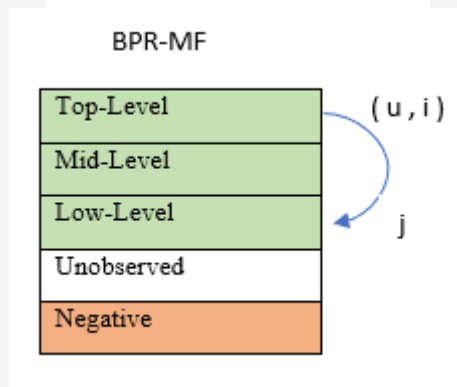
2.4. **Matrix Factorization (MF)**

MF merupakan teknik pemodelan data menggunakan *matrix*, yang berisikan faktor dari karakteristik barang dan penggunanya. Biasanya pada sistem rekomendasi memiliki beberapa tipe data yang dimasukkan kedalam *matrix*, dengan salah satu *matrix* direpresentasikan sebagai dimensi untuk user dan yang lainnya sebagai dimensi untuk barang yang diperhatikan. Kelebihan dari MF merupakan kemampuannya yang memungkinkan untuk menggabungkan beberapa informasi tambahan. Sehingga, jika *feedback* yang bersifat eksplisit tidak tersedia, sistem rekomendasi dapat menambahkan informasi menggunakan *feedback* yang bersifat implisit. Dengan kelebihan yang dimiliki MF, metode ini memiliki fleksibilitas dalam memodelkan beragam keadaan di dunia nyata[5].

2.5. **Bayesian Personalized Ranking – Matrix Factorization (BPR-MF)**

BPR-MF merupakan metode sistem rekomendasi yang menggunakan sifat BPR dengan model MF. Pada BPR-MF, data sample yang bersifat positif akan di urutkan berdasarkan tingkat ketertarikan seorang user terhadap suatu barang. Selain itu pada BPR-MF memisahkan data barang yang menjadi dua yakni barang yang tidak diketahui oleh pengguna, dan data barang yang tidak disukai oleh pengguna tersebut [6].

BPR-MF terbukti memiliki performa yang paling baik dalam menyusun tingkatan data yang berdasarkan dataset. Hal tersebut dikarenakan, BPR-MF tidak hanya berfokus pada pemodelan data, namun juga berfokus dalam mengoptimalkan kriteria data, sehingga BPR-MF dalam membagi data-data secara akurat [6]. Namun, BPR memiliki akurasi yang kurang dalam mengolah model berdimensi kecil [4].



Gambar 2. Contoh *sampling* data dengan pengurutan data positif berdasarkan tingkat ketertarikan pelanggan terhadap suatu barang, dan data negatif yang berupa barang yang tidak diketahui, serta barang yang tidak disukai [6].

Dari Gambar diatas kita akan bisa mendapatkan relasi antara dua *item* $i >_u j$, mengindikasikan bahwa pengguna u lebih tertarik barang i dibandingkan barang j. Untuk mengestimasi ketertarikan user terhadap sebuah barang dibandingkan dengan barang yang lain, formula ini dinamakan BPR-Opt dan didefinisikan sebagai [7] :

$$BPR - Opt := \sum_{(u,i,j) \in D_k} \ln \sigma(\hat{s}_{uij}) - \ln p(\theta) \tag{1}$$

Dimana $\hat{s}_{uij} := \hat{r}_{ui} - \hat{r}_{uj}$ dan $DK = \{(u, i, j) | i \in N(u) \& j \in \bar{N}(u)\}$ [8], dengan \hat{s}_{uij} merupakan nilai hasil perbandingan antara barang i atau j yang lebih sering dipilih oleh pengguna u, sedangkan \hat{r}_{ui} atau \hat{r}_{uj} merupakan prediksi *personalized score* untuk barang i atau j terhadap user u[8]. Simbol $p(\theta)$ dinyatakan sebagai parameter pembelajaran dari model, dan σ untuk fungsi logistik, didefinisikan sebagai : $\sigma(x) = 1/(1 + e^{-x})$

[8]. Berikut ini merupakan algoritma yang digunakan pada penelitian [8] untuk pembelajaran BPR pada sistem, dengan α sebagai rata-rata pembelajarannya :

```

Input      :  $D_k$ 
Output    : Learning parameter  $\Theta$ 
Inisialisasi  $\Theta$  dengan nilai acak
for count = 1,...,#Iter do
    draw  $(u, i, j)$  from  $D_k$ 
     $\hat{s}_{uij} := \hat{r}_{ui} - \hat{r}_{uj}$ 
     $\Theta \leftarrow \Theta + \alpha \left( \frac{e^{-\hat{s}_{uij}}}{1+e^{-\hat{s}_{uij}}} \cdot \frac{\partial}{\partial \Theta} \hat{s}_{uij} - \Lambda_{\Theta} \Theta \right)$ 
End

```

Pada penelitian [4] dibuktikan bahwa BPR-MF[4] lebih baik dibandingkan metode Cosine-kNN [4]. Hal tersebut disebabkan oleh, metode BPR-MF[4] mengklasifikasikan dua jenis prediksi yang berbeda yakni \hat{r}_{ui} dan \hat{r}_{uj} , sedangkan dalam Cosine-kNN hanya memprediksi satu kriteria yakni \hat{r}_{ui} . Sehingga, pada penelitian [4] dihasilkan performa BPR-MF yang lebih tinggi dibandingkan Cosine-kNN pada grafik Area Under the Curve(AUC).

2.6. Area Under Curved (AUC)

Area Under Curved (AUC) merupakan metode yang akan digunakan untuk mengevaluasi performansi pada sistem rekomendasi serta kualitas rekomendasi yang dihasilkan oleh BPR dengan pemodelan data menggunakan MF. AUC adalah prediksi *ranking* dari nilai positif yang terambil secara acak sebelum nilai negatif. Sistem akan menghitung nilai AUC berdasarkan hasil pengambilan data dari data yang telah dimodelkan menggunakan MF. Nilai AUC memiliki range dari nol sampai satu, dimana nilai satu merupakan nilai evaluasi terbaik, sedangkan nol sebaliknya. Hasil prediksi ini akan ditampilkan oleh sistem beserta banyaknya data yang diprediksi, berikut ini merupakan formula untuk perhitungan AUC :

$$AUC(u) := \frac{1}{|I_u^+||I \setminus I_u^+|} \sum_{i \in I_u^+} \sum_{j \in I \setminus I_u^+} \delta(\hat{x}_{uij} > 0) \quad (2)$$

dimana,

$$\delta(x > 0) = H(x) := \begin{cases} 1, & x > 0 \\ 0, & \text{else} \end{cases} \quad (3)$$

u : User

I : Indeks

I_u^+ : Indeks dari user u yang bernilai positif

δ : Fungsi Heaviside atau *non-differentiable loss*

\hat{x}_{uij} : Fungsi dengan nilai real yang didapat dari parameter model yang berisi relasi antara user u , *item* i dan *item* j

Dari formula tersebut sistem akan mengambil *item* (buku) yang bernilai positif atau yang nilainya lebih dari nol dari kumpulan *item* berdasarkan user, dan memasukkannya pada array. Setelah semua array user terpenuhi maka sistem akan menghitung rata-rata nilai AUC berdasarkan user tersebut dan menampilkannya pada layer. Semakin banyak user yang memiliki satu *item* dengan nilai positif maka, semakin besar nilai AUC user tersebut. AUC akan membandingkan matriks hasil prediksi dengan matriks pada pemodelan data dimana, jika nilai I_u^+ memiliki nilai positif atau negatif yang sama maka nilai AUC akan semakin mendekati satu, sedangkan jika sebuah nilai pada matriks prediksi dengan matriks pada model data berbeda maka nilai AUC akan lebih mendekati nilai nol. Fungsi utama penggunaan AUC yakni untuk, mengetahui seberapa besar ketepatan hasil prediksi BPR dalam meranking item yang dibandingkan dengan pemodelan yang dibuat oleh MF, sehingga ranking yang dihasilkan oleh BPR harus sesuai dengan data yang dimodelkan.

2.7. Summary

Pada penelitian [11][12] dilakukan penerapan metode yang berbeda untuk membangun sistem rekomendasi untuk merekomendasikan buku. Pada penelitian [11] dilakukan percobaan untuk membangun

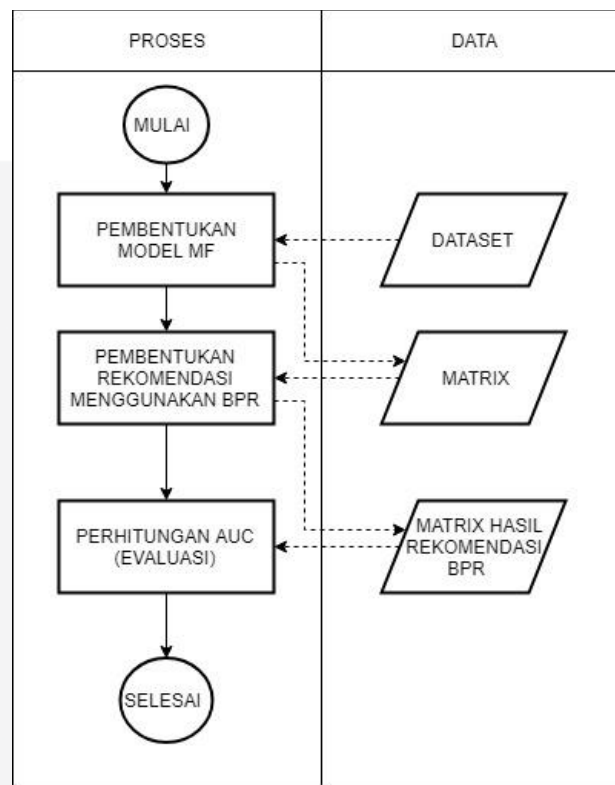
sistem rekomendasi menggunakan metode Collaborative Filtering (CF) yang dibantu dengan teknik Time Sequence Base(TSB) yakni teknik yang digunakan untuk mengetahui catatan waktu saat user berinteraksi dengan item yang hasilnya akan digunakan dalam sistem rekomendasi. Evaluasi pada penelitian tersebut menggunakan metode Top-N dimana hasil akurasi penelitian tersebut yakni 75% untuk merekomendasikan Top-20. Kelebihan dari penelitian tersebut yakni sistem rekomendasi dapat menghasilkan rekomendasi yang relevan terhadap personal seorang user, karena menggunakan metode TSB.

Pada penelitian [12] dilakukan penelitian terhadap metode Collaborative Filtering (CF) dan Association Mining untuk membangun sistem rekomendasi dalam merekomendasikan buku. Pada penelitian ini menggunakan metode evaluasi MAE dengan akurasi 70%. Dari kesimpulan yang didapat pada penelitian tersebut bahwa Item Based Collaborative Filtering dapat menghilangkan masalah terhadap sparsity data atau data item yang hanya dirating oleh beberapa user dan dapat menghasilkan rekomendasi yang baik.

3. Sistem yang Dibangun

3.1. Gambaran Umum Sistem

Berikut ini merupakan gambaran alur yang akan dilakukan oleh sistem pada tugas akhir ini :



Gambar 3. Gambaran umum sistem

Pada tugas akhir ini, sistem yang dibangun merupakan Sistem Rekomendasi Buku dengan menggunakan metode Bayesian Personalized Rankin dan pemodelan data menggunakan Matrix Factorization untuk merekomendasikan buku. Sistem yang dibangun bertujuan untuk mengatasi rekomendasi buku yang tidak akurat pada situs – situs buku online, yang dikarenakan sistem rekomendasi tersebut hanya merekomendasikan buku dengan rating buku yang tinggi atau tidak sesuai dengan preferensi pengguna.

Gambar 3 menjelaskan seluruh alur proses sistem rekomendasi yang dibangun pada tugas akhir ini yakni Pembentukan model MF, Pembentukan rekomendasi menggunakan BPR, dan perhitungan AUC(evaluasi). Pertama yakni Pembentukan Model MF adalah pemodelan dataset buku yang akan menjadi matrix user terhadap item seperti pada gambar 4. Model matrix [4] dimana, matrix pada sistem hanya akan bernilai satu dan nol, dimana nilai satu menyatakan bahwa data buku tersebut memiliki *feedback* positif atau diobservasi oleh *user*, sedangkan nilai 0 bagi data buku yang tidak diketahui oleh *user* atau data user dengan buku yang tidak terdaftar pada dataset. Pada tahap kedua yakni Pembentukan Rekomendasi Menggunakan BPR adalah memprediksi nilai nol pada sistem dengan menggunakan metode ranking yakni BPR, untuk memperkirakan data buku yang tidak diketahui oleh user, dan hal tersebut akan dilakukan untuk setiap *user*, pada tahap ini sistem akan menggunakan

sepuluh item dengan ranking tertinggi dari sebuah user untuk dibandingkan dengan user lainnya contohnya, ada tiga user yakni a, b, dan c dimana item x, y, dan z merupakan top item untuk user a dan b, sedangkan item x dan y merupakan top item untuk user c maka, sistem nantinya akan merekomendasikan item z untuk user c. Pada tahap terakhir yakni Perhitungan AUC(evaluasi) adalah metrik evaluasi yang digunakan oleh sistem untuk menilai rata-rata atau kualitas prediksi berdasarkan, jumlah kemungkinan terambilnya hasil prediksi rekomendasi yang bernilai positif dari hasil perhitungan BPR, dengan hasil negatif yang dihasilkan.

3.2. Dataset

Dataset yang digunakan dalam pengerjaan tugas akhir ini berasal dari *Goodreads* dan *Bookcrossing*, dataset tersebut merupakan hasil pengolahan data pada komunitas *Goodreads* [9], sedangkan *Bookcrossing* [13] merupakan dataset dari komunitas *Bookcrossing*, namun telah diolah oleh Cai-Nicolas Ziegler dan dataset tersebut diolah oleh Arthur Fortes[14], dengan menghilangkan user dan item yang memiliki interaksi kurang dari 20 dan 10. Pada tugas akhir ini, data yang dipakai merupakan data rating dari kedua dataset tersebut yang terdiri dari atribut *user_id*, *book_id*, dan *rating*, berikut ini contoh tabel data rating pada dataset *Goodreads* dan *Bookcrossing* yang dipakai pada tugas akhir ini:

Tabel 1. Contoh data rating *Goodreads*

<i>user_id</i>	<i>book_id</i>
1	258
2	4081
2	260
2	9296
2	2318

Dataset *Goodreads* berisikan 10.000 data buku dan 53.424 data user dan memiliki lima juta data rating, dengan tipe file yakni CSV(*Comma Separated Value*).

Tabel 2. Contoh data rating *Bookcrossing*

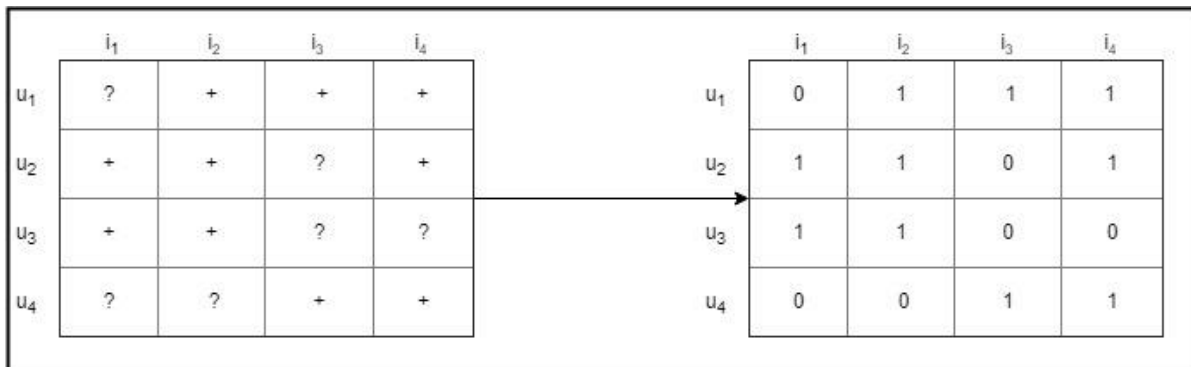
<i>user_id</i>	<i>book_id</i>
1	6264
1	4350
1	6252
1	202
1	6266

Dataset *Bookcrossing* berisikan 272.679 interaksi implisit dan eksplisit dari 2.946 user terhadap 17.384 buku. Dataset yang dipakai merupakan data rating, karena data rating menjelaskan bahwa user telah berinteraksi terhadap sebuah buku secara implisit maupun eksplisit. User yang memiliki riwayat dengan sebuah buku akan dimodelkan menjadi feedback positif pada sistem sedangkan, user yang tidak berinteraksi atau tidak ada pada data rating tersebut akan bersifat feedback negatif pada sistem.

3.3. Pemodelan Data Matrix Factorization

Pada tahap ini, sistem memodelkan dataset yang ada menjadi matrix. Pemodelan ini menggunakan metode *Matrix Factorization*. Dimana baris pada *matrix* akan menggunakan atribut *user_id*, sedangkan kolom *matrix* akan menggunakan atribut *book_id*, berikut ini contoh gambar matriks yang akan dibuat oleh sistem

dimana U merupakan *user* dan i merupakan *item* :

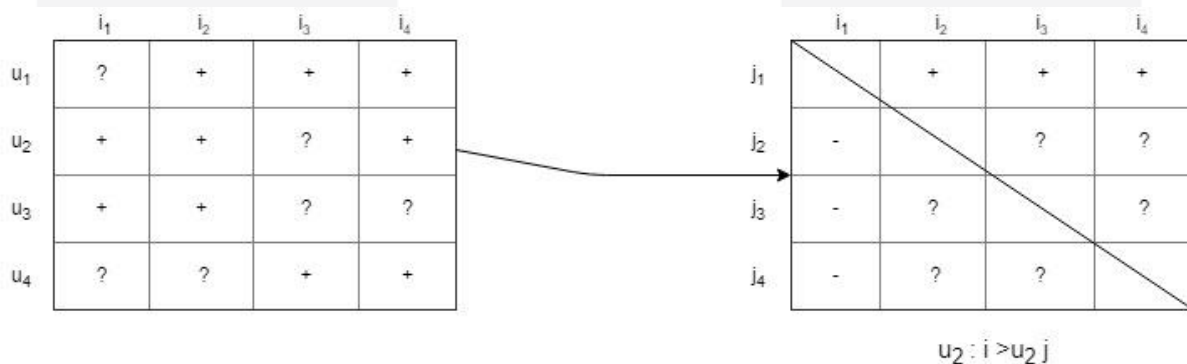


Gambar 4. Model matrix [4]

Dari gambar diatas, matrix sebelah kanan merupakan pemodelan awal data buku dari dataset, dimana matrix tersebut hanya berisikan *feedback* positif atau buku yang telah diobservasi oleh *user*. Sedangkan, buku yang tidak diobservasi belum ditentukan nilainya. Sehingga pemodelan pada sistem akan berupa matrix yang sebelah kanan, dimana nilai satu merupakan *feedback* positif atau buku yang telah diobservasi oleh *user*, dan nilai nol merupakan buku yang tidak diobservasi *user* atau tidak diketahui oleh *user*.

3.4. Model Rekomendasi Bayesian Personalized Ranking

Pada tahap ini sistem akan mengisi data yang kosong dari pemodelan data MF. Dibawah ini contoh model matrix yang sistem buat untuk mengisi data kosong pada pemodelan data MF dengan prediksi nilai ranking *item* dari hasil perhitungan pada Persamaan 4 untuk setiap U terhadap *item* I dan j, berikut ini contoh model matrix yang dibuat oleh sistem :



Gambar 5. Model matrix dengan hasil ranking

Dapat dilihat dari gambar diatas, matriks sebelah kiri merupakan pemodelan awal dari data buku. Sedangkan, matriks sebelah kanan merupakan pecahan matriks yang lebih spesifik untuk user, dari gambar diatas matriks sebelah kiri merupakan matriks spesifik untuk u_2 . Dari matriks sebelah kanan dapat dilihat bahwa nilai positif menandakan bahwa u_2 lebih memilih *item* i dibandingkan *item* j, sedangkan jika kedua *item* tersebut memiliki nilai positif atau sama – sama tidak diobservasi oleh *user* maka, nilai pada matriks blum diketahui sehingga dibutuhkan metode ranking untuk mengetahui nilai tersebut.

$$BPR - Opt := \sum_{(u,i,j) \in D_k} \ln \sigma(\hat{s}_{uij}) - \ln p(\theta) \tag{4}$$

$$\sigma(x) := \frac{1}{1+e^{-x}} \tag{5}$$

\hat{s}_{uij} : Nilai hasil pembandingan interest user u terhadap barang i atau j

\hat{r}_{ui} : Nilai prediksi *personalized score* untuk barang i terhadap user u

\hat{r}_{uj} : Nilai prediksi *personalized score* untuk barang j terhadap user u

$p(\Theta)$: Parameter pembelajaran dari model

σ : fungsi logistik

Dimana $p(\Theta)$ merupakan normal distribusi dengan rata-rata mendekati nol dengan nilai rata-rata yakni nilai join variability antara dua random item pada matrix, menggunakan perhitungan variance-covariance matrix sebagai berikut,

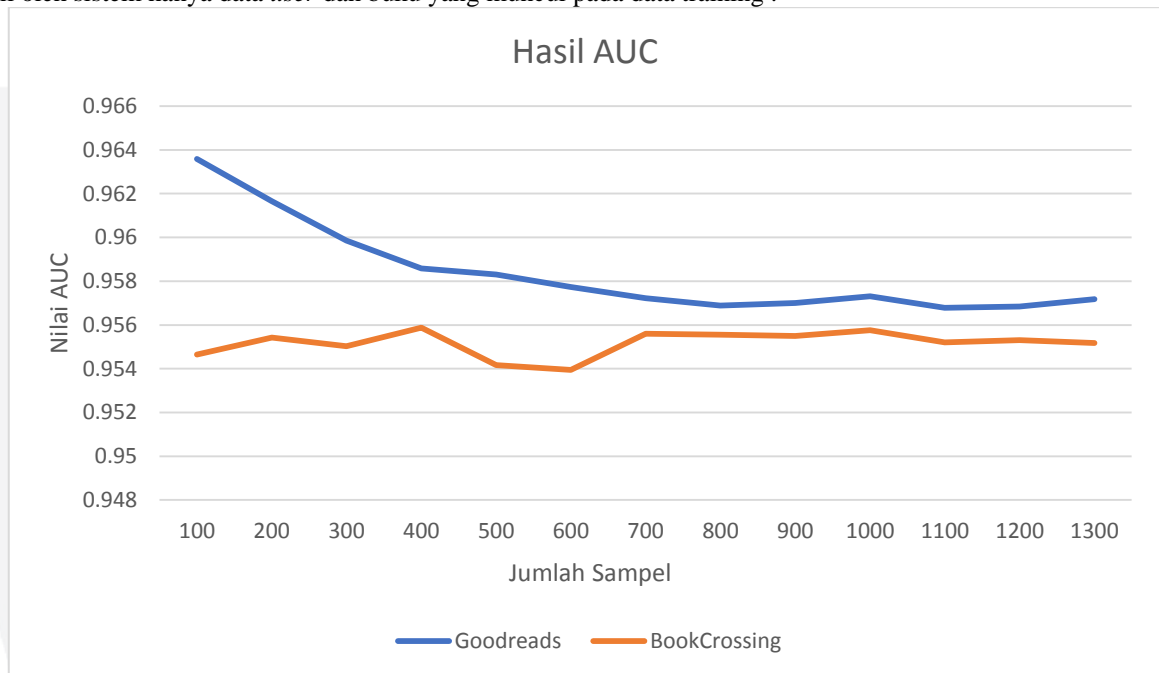
$$\Sigma_{ij} = cov(X_i, X_j) = E[(X_i - \mu_i)(X_j - \mu_j)] \tag{6}$$

dimana Σ merupakan matrix yang memiliki nilai i, j, dan E merupakan operator untuk menghitung rata-rata hasil dari covariance tersebut.

4. Evaluasi

4.1. Hasil Pengujian

Pada tugas akhir ini dataset yang digunakan merupakan dataset *Goodreads* dan *Bookcrossing*, dengan nilai $AUC(u) = 0.95718$ untuk dataset *Goodreads* dan $AUC(u) = 0.95518$ untuk dataset *Bookcrossing*. Berikut ini merupakan diagram hasil perhitungan AUC berdasarkan sample yang diambil oleh sistem, sampel yang diambil oleh sistem hanya data *user* dan buku yang muncul pada data training :



Gambar 5. Diagram AUC

Diagram AUC diatas menjelaskan nilai AUC yang didapat oleh sistem untuk stiap 1000 sample. AUC hanya memiliki nilai. Berikut ini merupakan tabel yang berisi sebagian mapping dari hasil matrix BPR yang dihasilkan oleh sistem :

Tabel 3. Contoh hasil mapping pada sistem

Indeks menurut id user (I_u)	Indeks menurut id buku (I_i)	User_id (u)	Book_id (i)	Hasil mapping (I_u, I_i)
0	0	1	258	(0,0)
1	1	2	4081	(1,1)
1	2	2	260	(1,2)
1	3	2	9296	(1,3)
1	4	2	2318	(1,4)

Tabel diatas menjelaskan mengenai mapping atau hubungan yang terjadi antara user dan item, pada sistem setiap id_user dan id_buku akan disimpan pada array dan dapat dipanggil berdasarkan indeks array. Hasil mapping merupakan data item yang diobservasi oleh user, yang terdiri dari indeks $user$ yang terhubung dengan indeks $item$.

4.2. Analisis Hasil Pengujian

Berdasarkan hasil pengujian pada gambar 5, pengujian tersebut dilakukan pada dataset *Goodreads* dan *Bookcrossing*. Dari hasil AUC yakni dengan dataset *Goodreads* dengan nilai $AUC(u) = 0.95718$ dan $AUC(u) = 0.95518$ untuk dataset *Bookcrossing* dengan data sampel yang dipakai yakni sebanyak 1200, dapat dinyatakan bahwa pemodelan data menggunakan Matrix Factorization dengan dataset *Goodreads* dapat menghasilkan rekomendasi yang lebih baik dibandingkan *bookcrossing*, hal ini dapat disebabkan karena jumlah data rating *goodreads* lebih banyak dibandingkan *bookcrossing*, karena semakin banyak data rating yang diprediksi pada BPR maka prediksi ranking sejumlah item untuk seorang user akan semakin akurat.

5. Kesimpulan

Berdasarkan hasil analisis dan pengujian pada tugas akhir ini dapat disimpulkan bahwa, Bayesian Personalized *Ranking* dengan menggunakan model Matrix Factorization dapat digunakan untuk merekomendasikan buku, dengan data sebanyak satu juta. Dari hasil yang didapat yakni $AUC = 0.95718$ dan $AUC = 0.95518$ menyatakan bahwa, metode tersebut dengan pemodelan data menggunakan MF menghasilkan prediksi rekomendasi yang baik, karena nilai AUC yang didapatkan mendekati nilai satu.

Daftar Pustaka

- [1] Kenta Kawase, Bang Hai Le, dan Ruck Thawonmas, "Collaborative Filtering for Recommendation of Areas in Virtual Worlds," Network and Systems Support for Games (NetGames), 2014 13th Annual Workshop on, 5 Desember 2014
- [2] Anuranjan Kumar, Sahil Gupta, S. K Singh, dan K. K. Shukla," Comparison of Various Metrics Used in Collaborative Filtering for Recommendation System ," Contemporary Computing (IC3), 2015 Eighth International Conference on, 22 Aug. 2015
- [3] Rohit Parimi, Doina Caragea, "Community Detection on Large Graph Datasets for Recommender Systems," Data Mining Workshop (ICDMW), 2014 IEEE International Conference on, 14 Desember 2014
- [4] Steffen Rendle, Christoph Freudenthaler, Zeno Gantner and Lars Schmidt-Thieme, "BPR: Bayesian Personalized Ranking from Implicit *Feedback*," UAI '09 Proceedings of the Twenty-Fifth Conference on Uncertainty in Artificial Intelligence, June 18 - 21, 2009
- [5] Yehuda Koren, Robert Bell and Chris Volinsky, "Matrix Factorization Techniques for Recommender Systems," Journal Computer Volume 42 Issue 8, August 2009
- [6] Babak Loni, Roberto Pagano, Martha Larson, Alan Hanjalic, "Bayesian Personalized Ranking with Multi-Channel User *Feedback*," RecSys '16 Proceedings of the 10th ACM Conference on Recommender Systems, 15 – 19 September 2016
- [7] Nursultan Kurmashov, Konstantin Latuta, "Online Book Recommendation System," Electronics Computer and Computation (ICECCO), 2015 Twelve International Conference on, 27 – 30 September 2015
- [8] Marcelo G. Manzato, Marcos A. Domingues, Solange O. Rezende, "Optimizing Personalized Ranking in Recommender Systems with Metadata Awareness," Web Intelligence (WI) and Intelligent Agent Technologies (IAT), 2014 IEEE/WIC/ACM International Joint Conferences on, 11-14 Agustus 2014
- [9] Ritesh Kumar, Guggila Bhanodai, Rajendra Pamula, "Social book search: Reranking based on document and query expansion with keyword Filtering," Advanced Computing and Communication Systems (ICACCS), 2017 4th International Conference on, 6-7 Januari 2017
- [10] Ng Annalyn, Maarten W. Bos, Leonid Sigal, and Boyang Li, "Predicting Personality from Book Preferences with User-Generated Content Labels," IEEE Transactions on Affective Computing, 23 February 2018
- [11] Fuli Zhang," A Personalized Time-Sequence-based Book Recommendation Algorithm for Digital Libraries," IEEE Access (Volume: 4), 09 May 2016
- [12] Suhasini Parvatikar, Dr. Bharti Joshi," Online Book Recommendation System by using Collaborative filtering and Association Mining," 2015 IEEE International Conference on Computational Intelligence and Computing Research (ICCIC), 21 March 2016
- [13] Adli Ihsan Hariadi, Dade Nurjanah," Hybrid Attribute and Personality based Recommender System for Book Recommendation," 2017 International Conference on Data and Software Engineering (ICoDSE), 1-2 Nov. 2017
- [14] Arthur F. da Costa, Marcelo G. Manzato," Multimodal Interactions in Recommender Systems: An Ensembling Approach," 2014 Brazilian Conference on Intelligent Systems, 18-22 Oct. 2014