

Handling Sparsity to Improve Quality Recommendation on Collaborative Filtering

Seli Puri A

Department of Informatics, Muhammadiyah University of Lampung, Lampung, Indonesia
E-mail: selipuria@gmail.com

Nirmala Handayani

Department of Informatics, Muhammadiyah University of Lampung, Lampung, Indonesia
E-mail: nirmalahandayani@mail.uml.ac.id

*Corresponding Author

Ahmad Rofi'i

Department of Informatics, Muhammadiyah University of Lampung, Lampung, Indonesia
E-mail: ahmadrofii@mail.uml.ac.id

Received: 20 November 2024; Accepted: 25 November, 2024; Published: 30 November, 2024

Abstract: The development of e-commerce platforms in this century has been very rapid and currently there are many e-commerce sites in Indonesia that sell various kinds of products and services. A recommendation system is a system for filtering, sorting items and information that takes preferences from user behavior, user profiles or opinions from the user community to assist individuals in identifying interesting and potentially high-potential content to select, purchase or use. In this study, data that continues to grow requires a recommendation as an alternative to refinding the items needed. This is because the resulting recommendations have better quality. Until this research tries to analyze sparsity handling. From the results of the evaluation, the conclusion that can be drawn is that the evaluation process results in a higher ranking of 1% Borda-count. Based on the results of the significance test, the Borda Algorithm runs around 28,000 times faster. In addition, the imputation stages carried out also affect the predicted ranking value. Meanwhile, based on accuracy metrics, the proposed method obtains 2 times higher Coverage values, 19% higher NDCG values.

Keywords: Borda, Imputation, Collaborative Filtering, Ranking, Sparsity.

1. Introduction

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

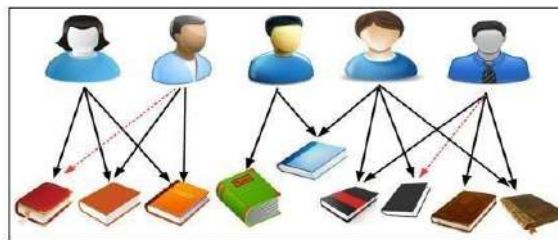
Sparsity merupakan suatu kondisi yang terjadi akibat terdapat banyak kekosongan data rating pada item, sehingga menyebabkan nilai rating pengguna terhadap item cenderung renggang. Hal ini disebabkan karena pengguna cenderung hanya memberikan nilai ratingnya terhadap beberapa item saja. Masalah sparsity merupakan kendala yang ada pada sistem rekomendasi dengan metode Collaborative Filtering karena pada metode Collaborative Filtering rating merupakan hal yang paling penting dalam memberikan rekomendasi. Semakin banyak rating yang terkumpul maka semakin baik hasil rekomendasi yang diberikan. Semakin tinggi sparsity data rating maka akan memberikan akurasi yang buruk (akurasi yang rendah) terhadap hasil rekomendasi (Xue, 2009) dalam (Sanjung, 2011). Jumlah user dan jumlah item yang besar akan sangat membantu proses perekomendasi, namun tidak mungkin semua user akan aktif menggunakan sistem dan memberikan penilaian pada setiap item.

Imputasi adalah proses pengisian nilai kosong terhadap data menggunakan metode tertentu [9]. Salah satu metode dasar dalam melakukan pengisian data adalah menggunakan mean. Nilai mean menunjukkan kecenderungan besarnya suatu nilai dalam kumpulan data, dan dapat digunakan untuk menggantikan kekosongan suatu nilai pada kumpulan data seperti pada [9]. Penggunaan mean pada Imputation-Boosted NBCF adalah dengan membentuk matriks preferensi yang berisi rating user terhadap genre. Nilai pada matriks preferensi merupakan nilai mean dari kumpulan rating user terhadap genre tertentu, menggunakan persamaan (1).

Sistem rekomendasi mempunyai tujuan untuk menghasilkan rekomendasi yang bermakna bagi penggunanya mengenai hal-hal yang mungkin menarik bagi mereka. Beberapa metode populer yang digunakan dalam membuat rekomendasi adalah pemfilteran berbasis konten, pemfilteran kolaboratif, dan hibrid. Pemfilteran berbasis konten memanfaatkan interaksi antara konten item dan profil pengguna yang mencakup konten item seperti genre, penulis, dll. Pemfilteran kolaboratif (CF) bekerja dengan membangun database preferensi konsumen terhadap suatu item. Seperti promosi dari mulut ke mulut, pemfilteran kolaboratif memberikan prediksi peringkat dan rekomendasi pribadi berdasarkan apa yang disukai pengguna lain yang memiliki selera yang sama [6]. Sedangkan hybrid merupakan kombinasi dari content based filtering dan Collaborative Filtering.

2. Research Method

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

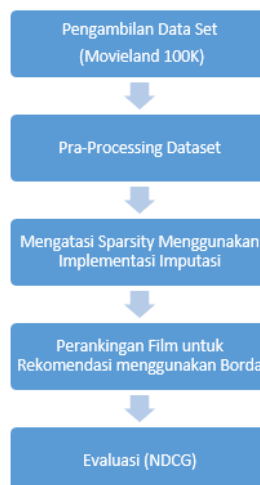


Gambar 1. Collaborative flittering

Adapun penjelasan dari masing-masing tahap dalam metode *Collaborative flittering* adalah sebagai berikut. Perencanaan, pada tahap ini perancangan perangkat lunak sistem informasi geografis yaitu perencanaan yang bertujuan untuk membuat skema sistem yang akan dibuat yaitu sistem informasi geografis yang menunjukkan kualitas udara di kampus unama Jambi.

a. Diagram Alur Penelitian

Alur penelitian yang digunakan dalam penelitian ini terdiri dari 5 tahap yang digambarkan pada Gambar 2 Pada tahapan ini memaparkan kegiatan yang dilakukan pada pengembangan metode Collaborative Filtering dengan penjelasan sebagai berikut :



Gambar 2. Diagram Alur Penelitian

- b. Pengambilan Data Set (MovieLens 100k)
Tahapan ini dilakukan dengan mengumpulkan dataset movieLens 100k sebagai bahan evaluasi terhadap metode yang dikembangkan. Dataset yang digunakan pada penelitian ini adalah data Movie yang diakses dari website GroupLens (<https://grouplens.org/datasets/movielens/>). Dataset yang digunakan adalah MovieLens 100K.
- c. Pra- prosesi Dataset
Bertujuan untuk eliminasi data yang tidak digunakan dalam penelitian ini. Adapun Dataset MovieLens merupakan data yang valid dan terupdate secara berkala, pra-pengelolaan dataset tetap perlu dilakukan untuk meningkatkan performa sistem rekomendasi.
- d. Mengatasi Sparsity menggunakan Implementasi Imputasi
Proses mengatasi Sparsity merupakan kendala yang ada pada sistem rekomendasi dengan metode collaborative filtering karena pada metode collaborative filtering merupakan hal yang paling penting dalam pemberian rekomendasi. Hal tersebut Penggunaan Implementasi Imputasi merupakan suatu teknik dalam aljabar linear yang memiliki banyak fungsi dalam pengolahan Sparsity dalam dataset.
- e. Perengkingan Film untuk rekomendasi menggunakan Borda
Proses perengkingan film ini bertujuan untuk merekomendasikan kepada pengguna tentang produk tersebut sehingga pengguna mendapatkan rekomendasi sesuai preferensi user.
- f. Evaluasi (NDCG)
Langkah selanjutnya adalah melakukan evaluasi hasil penelitian yang dilakukan dengan beberapa teknik evaluasi.

2.1 Alat Bantu Penelitian

- 1. Perangkat Keras (hardware)
 - o Laptop Asus, Memori 320GB, Corei3
 - o Flashdisk Sandisk 8GB
 - o Printer Canon IP 2700
 - o Modem Portebel Smart Fren – M5
- 1. Perangkat Lunak (software)
 - o OS (Windows 7 Ultimate 64-Bit)
 - o Microsoft Word 2007
 - o Microsoft Power Point 2007
 - o Web Browser (Mozilla Firefox)

Penelitian ini menggunakan dataset yang umum dan banyak digunakan dalam sistem rekomendasi collaborative filtering yaitu dataset movie dari MovieLens [17]. Dataset tersebut dapat diakses di website GroupLens (<https://grouplens.org/datasets/movielens/>). Dataset MovieLens merupakan dataset open source yang dapat digunakan secara bebas namun harus mengikuti aturan yang telah dibuat oleh GroupLens, yang merupakan laboratorium penelitian yang berada di University of Minnesota, USA. Dataset yang digunakan dalam eksperimen ini yaitu MovieLens 100K. Dataset 100K memiliki sekitar 1700 movie, 1000 pengguna dan 100,000 rating. Dataset tersebut memiliki karakteristik umum yaitu informasi demografis pengguna (user id, gender, age, occupation, dan zipcode). Dataset ini juga memiliki 19 genre seperti pada Tabel 3.1. Selain itu setiap pengguna minimal memberikan rating pada 20 movie. Dataset tersebut mengandung sparsity 95,8% [18], hal tersebut disebabkan karena banyak pengguna tidak memberikan rating ke movie.

3. Results and Discussion

3.1. Collaborative Filtering

- Metode Ranking Based *Collaborative Filtering* menghasilkan performa yang lebih baik dalam mengatasi masalah sparsity karena dataset yang bersifat renggang dibandingkan dengan Rating Based Collaborative Filtering. metode ini masih menghadapi masalah sparsity. Penelitian ini mengombinasikan Imputation dengan Ranking Based Collaborative Filtering untuk mereduksi dimensi dataset (matriks rating) untuk mengatasi masalah sparsity yang dihadapi metode *Ranking Based Collaborative Filtering*.
- Metode Ranking Based Collaborative Filtering menghasilkan performa yang lebih baik dalam mengatasi masalah sparsity karena dataset yang bersifat renggang dibandingkan dengan Rating Based Collaborative Filtering. metode ini masih menghadapi masalah sparsity. Penelitian ini mengombinasikan Imputation dengan Ranking Based Collaborative Filtering untuk mereduksi dimensi dataset (matriks rating) untuk mengatasi masalah sparsity yang dihadapi metode Ranking Based Collaborative Filtering.

3.2 Mengatasi Sparsity Menggunakan Implementasi *Imputasi (Imputation)*

Untuk mengatasi kekosongan rating menggunakan Implementasi *Imputasi (Imputation)*. Hasil dari *Imputasi* ini adalah merupakan implemtasi yang menangani pada rating yang memiliki kekosongan. Dimana nilai *imputasi* menyimpan informasi yang sangat penting tentang data yaitu data yang berkontribusi paling besar terhadap variasi data secara keseluruhan.

Tabel 1. Tabel dengan sparticy

		1	2	3	4	5	6	7	8	9	10	934
1	- Toy Story (1995)	0	0	0	1	1	1	0	0	0	0	1
2	- GoldenEye (1995)	0	1	1	0	0	0	0	0	0	0	0
3	- Four Rooms (1995)	0	0	0	0	0	0	0	0	0	0	3
4	- Get Shorty (1995)	0	1	0	0	0	1	0	0	1	0	0
5	- Copycat (1995)	0	0	0	0	0	0	1	0	1	0	3
1683	- Scream of Stone	0	0	0	0	0	0	0	0	1	0	4

3.2. Implementasi Metode Borda

Ide dari metode borda adalah mengharuskan para pemilih memberikan ranking kepada tiap kandidat, serta memberikan nilai untuk tiap-tiap peringkat. Proses yang dilakukan dengan melihat sejauh mana suatu produk lebih disukai oleh pengguna. Misalnya pada pengguna ke-1 = $A > B > C > D$. Hasil akhir metode Borda diperoleh dari menjumlahkan hasil perkalian antara jumlah pengguna dengan poinnya untuk setiap produk. Berdasarkan hasil dari proses Borda nilai terbesar dimiliki oleh film B sehingga saat di ranking akan dihasilkan urutan film yaitu B A C D. Berikut ini adalah proses pengkodean untuk menghasilkan ranking pada metode borda :

```
clear all;clc;
dataset = load('movies.mat');
dataRating = load('rating.mat');
data = dataset.movie;
rating = dataRating.rating;
%disp(data(1:1003, 1:3))

dataTrainRating = rating(1:2000, 2:2);

%disp(dataTrainRating)

[BordaWinner, candidate, bordaPoint] = BordaCount(dataTrainRating);
disp('Movie ID');
disp(BordaWinner)
disp('-----');
dataRelevance = [];

for i=1:10
    aa = data(i,i, 1:1);
    dataRelevance(i) = (candidate(1));
end

disp('NDCG');
a = ndcg(candidate, dataRelevance);
disp(a)
disp('-----');
disp('10 kandidat');
disp(candidate);

Movie ID
    1464

-----
NDCG
    0.9977
```

Gambar 3. Pengkodean Borda

Hasil yang diperoleh dengan metode borda dapat dilihat pada tabel 2.

Tabel 2. Data yang sudah melewati metode borda

MoviId	Film	TotalRank
1673	- Mirage (1995)	1673
1674	- Mamma Roma (1962)	1670
1675	- Sunchaser, The (1996)	1671
1777	War at Home, The (1996)	1672
1678	Sweet Nothing (1995)	1673
1779	Mat' i syn (1997)	1674
1690	B. Monkey (1998)	1679

3.3. Evaluasi

Tahapan evaluasi dilakukan dengan membandingkan hasil dari kinerja dan hasil rekomendasi dari Borda dan memanfaatkan normalized discounted cumulative gain (NDCG) matriks. Dari hasil evaluasi, kesimpulan yang didapat adalah bahwa proses evaluasi menghasilkan ranking lebih tinggi 1% Borda-count. Berdasarkan hasil uji signifikansi menunjukkan Algoritme Borda berjalan sekitar 28.000 kali lebih cepat. Selain itu tahapan Imputasi yang dilakukan juga berpengaruh terhadap nilai prediksi ranking. Sementara itu, berdasarkan metrik-metrik akurasi, metode yang diusulkan memperoleh nilai Coverage 2 kali lebih tinggi, nilai NDCG 19% lebih tinggi.

Peningkatan kualitas rekomendasi yang terjadi pada metode Borda dipengaruhi oleh pemanfaatan lebih lanjut data rating dalam proses pengolahan. Hal ini menyebabkan produk yang banyak pendapatan rating dari pengguna akan difilter. Selain itu dengan adanya bobot yang memanfaatkan rating akan berpengaruh terhadap produk yang memiliki rating besar lebih berpeluang menempati ranking yang lebih baik.

Tabel 3. Tabel hasil evaluasi

	Sparsity	Dengan Borda
NDCG 1	0,12	0,99
NDCG 2	0,16	1,00
NDCG 3	0,19	0,99
NDCG 4	0,20	0,87
NDCG 5	0,23	1,00

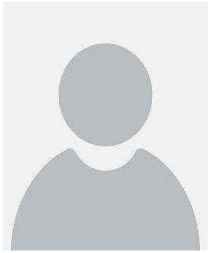
4. Conclusion

Penelitian ini mengusulkan metode Borda yang memanfaatkan Imputasi sebagai penanganan Sparsity. Imputasi dapat memberikan Rating yang semula "0" menjadi terisis sehingga pada perhitungan borda dapat menentukan ranking. Selanjutnya metode borda menghasilkan rekomendasi dengan cara mengalikan setiap bobot pada Film tersebut. Hasil pengujian menunjukkan bahwa pemanfaatan data rating dalam proses perankingan memberikan dampak terhadap peningkatan kualitas rekomendasi. Hal tersebut ditunjukkan dengan adanya peningkatan NDCG. sementara itu, berdasarkan metrik-metrik akurasi, metode yang diusulkan memperoleh nilai Coverage 2 kali lebih tinggi, nilai NDCG 19% lebih tinggi.

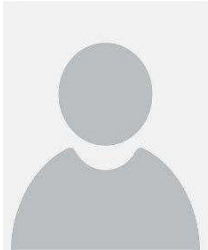
References

- [1] I. S. Wahyudi, "Big data analytic untuk pembuatan rekomendasi koleksi film personal menggunakan Mlib. Apache Spark," *Berk. Ilmu Perpust. dan Inf.*, vol. 14, no. 1, p. 11, 2018, doi: 10.22146/bip.32208.
- [2] H. Zarzour *et al.*, "Algoritma Rekomendasi Pemfilteran Kolaboratif Baru Berdasarkan Pengurangan Dimensi dan Teknik Pengelompokan," pp. 102–106, 2018.
- [3] M. Robbani, R. D. Ramadhani, and A. E. Amalia, "Analisa Algoritma Cosine Similarity dengan Pearson Correlation pada Metode Item-based Collaborative Filtering dengan Menggunakan Dataset Movielens," *Conf. Electr. Eng. Telemat. Ind. Technol. Creat. Media 2018*, pp. 88–92, 2018.
- [4] I. Yoshua and H. Bunyamin, "Pengimplementasian Sistem Rekomendasi Musik Dengan Metode Collaborative Filtering," *J. Strateg. ...*, vol. 3, pp. 1–16, 2021, [Online]. Available: <https://www.strategi.it.maranatha.edu/index.php/strategi/article/view/220>.
- [5] R. N. Sakinah, *Analisis Sistem Rekomendasi Data Rating Airbnb Menggunakan Inisialisasi Non-Negative Double Singular Value Decomposition Pada Metode Non-Negative Matrix Factorization*. 2020.
- [6] I. S. Wahyudi, *Mesin Rekomendasi Film Menggunakan Metode Kemiripan Genre Berbasis Collaborative Filtering*. 2017.
- [7] D. N. Ilham and S. Mulyana, "Sistem Pendukung Keputusan Kelompok Pemilihan Tempat PKL mahasiswa dengan Menggunakan Metode AHP dan Borda," *IJCCS (Indonesian J. Comput. Cybern. Syst.*, vol. 11, no. 1, p. 55, 2017, doi: 10.22146/ijccs.16595.
- [8] R. Siringoringo, J. Jamaluddin, and G. Lumbantoruan, "Sistem Perekomendasi Dengan Singular Value Decomposition Dan Teknik Similaritas Pearson Correlation," *Method. J. Tek. Inform. dan Sist. Inf.*, vol. 7, no. 1, pp. 19–24, 2021, doi: 10.46880/mtk.v7i1.257.
- [9] R. Rismala, "Analisis dan Implementasi Imputation-Boosted Neighborhood-Based Collaborative Filtering Menggunakan Genre Film," *Indones. J. Comput.*, vol. 2, no. 1, p. 31, 2017, doi: 10.21108/indojc.2017.2.1.50.
- [10] D. Statistika, F. Matematika, D. A. N. Ilmu, and P. Alam, "PENGELOMPOKAN DOSEN ITS BERBASIS," 2017.
- [11] A. Hapsery, "Regresi Kuantil Berbasis Model Rekursif dan Estimasi Sparsity Untuk Analisis Publikasi Dosen ITS di Scopus," *Thesis*, p. 143, 2017, [Online]. Available: <http://repository.its.ac.id/3044/>.
- [12] F. Arifin, E. Darwiyanto, and ..., "Perangkingan Fatwa Islam Dengan Metode Latent Semantic Indexing Menggunakan Teknik Singular Value Decomposition (svd)," *eProceedings ...*, vol. 6, no. 2, pp. 9414–9425, 2019, [Online]. Available: <https://librarye proceeding.telkomuniversity.ac.id/index.php/engineering/article/viewFile/9976/9832>.
- [13] S. Sulaiman and S. Agoes, "Analisis Reduksi Data Citra Menggunakan Metode Dekomposisi Nilai Singular," *Citee*, pp. 21–25, 2017, [Online]. Available: <https://docplayer.info/51307775-Analisis-reduksi-data-citra-menggunakan-metode-dekomposisi-nilai-singular.html%0Ahttps://docplayer.info/38429178-Watermarking-dengan-metode-dekomposisi-nilai-singular-pada-citra-digital.html>.
- [14] P. Geofisika, A. F. Sasti, and A. F. Sasti, "Komputasi Geofisika 1 : Singular Value Decomposition untuk Matriks 5x5," 2018, doi: 10.6084/m9.figshare.5946976.
- [15] F. Aryani, D. Yulianti, and J. Matematika, "Aplikasi Metode Singular Value Decomposition (SVD) Pada Sistem Persamaan Linier Kompleks," *J. Sains dan Teknol. Ind. UIN Sultan Syarif Kasim Riau*, vol. 10, no. 1, pp. 67–76, 2012.
- [16] F. Septian, "Penerapan Metode Singular Value Decomposition (SVD) pada Aplikasi Pengenalan Wajah Manusia," *J. Inform. Univ. Pamulang*, vol. 1, no. 2, pp. 54–59, 2016.
- [17] B. A. B. Ii and T. Pustaka, "BAB II Tinjauan Pustaka BAB II TINJAUAN PUSTAKA 2.1," vol. 1, pp. 1–64, 2002

Authors' Profiles



Seli Puri Andini merupakan dosen di Universitas Muhammadiyah Lampung, Fakultas Teknik Jurusan Informatika. Merupakan lulusan Starata 1 Universitas Mitra Indonesia yang meneruskan ke Strata 2 di IIB Darmajaya jurusan Magister Teknik Informatika



Nirmala Handayani, merupakan dosen di Universitas Muhammadiyah Lampung, Fakultas Teknik Jurusan Informatika. Merupakan lulusan Starata 1 IIB Darmajaya yang kemudian meneruskan ke Strata 2 IIB Darmajaya jurusan Magister Teknik Informatika..



Ahmad Rofi'i, merupakan dosen di Universitas Muhammadiyah Lampung, Fakultas Teknik Jurusan Informatika. Merupakan lulusan Starata 1 IIB Darmajaya yang kemudian meneruskan ke Strata 2 IIB Darmajaya jurusan Magister Teknik Informatika.