

Optimasi Pearson Correlation untuk Sistem Rekomendasi menggunakan Algoritma Firefly

Melany Mustika Dewi

Fakultas Ilmu Komputer, Universitas Amikom
Yogyakarta, Indonesia

e-mail: melany@amikom.ac.id

Informasi Artikel Diterima: 24-02-2021 Direvisi: 28-11-2021 Disetujui: 03-12-2021

Abstrak

Saat ini dunia digital berkembang untuk menyajikan informasi yang dibutuhkan dalam kehidupan. Informasi yang tepat bagi pengguna dengan banyak pertimbangan variabel adalah suatu permasalahan urgensi untuk saat ini. Dalam dunia digital film, menyajikan rekomendasi film yang sesuai dengan user merupakan suatu permasalahan. Solusi yang tepat untuk masalah tersebut adalah dengan Sistem Rekomendasi. Penelitian ini menggunakan teknik rekomendasi *Collaborative Filtering* dengan metode *Pearson Correlation* dan algoritma *Firefly* dalam menentukan rekomendasi film. Kelemahan dari Teknik *Collaborative Filtering* adalah munculnya *Sparsity* atau kekosongan data. Untuk mengatasi kelemahan ini peneliti akan melakukan penghapusan dari data yang kosong. Selain itu peneliti menggunakan metode *Pearson Correlation* untuk mencari nilai kemiripan antar user dan menggunakan Algoritma *Firefly* untuk menentukan film yang paling sesuai dengan user. Hasil dari penelitian ini menghasilkan *Mean Absolute Error* atau nilai kemungkinan salah sebesar 0,82 dan nilai akurasi sebesar 79,295%. Penggunaan metode *Pearson Correlation* tidak menghasilkan optimasi dalam penentuan kemiripan user yaitu mendapatkan error lebih rendah 0,21 yaitu sebesar 0,61 dan akurasi sebesar 83,83% atau lebih tinggi 4,535.

Kata Kunci: Pearson Correlation; Algoritma Firefly; Rekomendasi Film

Abstract

Currently the digital world develops to provide information needed in life. Appropriate information for users with many variable considerations is a matter of urgency for now. In the world of digital film, presenting film recommendations that suit the user is a problem. The perfect solution to this problem is the Recommendation System. This study uses the Collaborative Filtering recommendation technique with the Pearson Correlation method and the Firefly algorithm in determining film recommendations. The weakness of the Collaborative Filtering Technique is the appearance of sparsity or data gaps. To overcome this weakness, the researcher will erase the blank data. In addition, researchers used the Pearson Correlation method to find the similarity value between users and used the Firefly Algorithm to determine the most suitable film for the user. The results of this study produce a Mean Absolute Error or error probability value of 0.82 and an accuracy value of 79.295%. The use of the Pearson Correlation method does not result in optimization in determining the similarity of users, namely getting a lower error of 0.21, which is 0.61 and an accuracy of 83.83% or higher is 4.535.

Keywords: *Pearson Correlation; Firefly Algorithm; Recommended Films*

1. Pendahuluan

Film adalah salah satu karya digital sebagai hiburan untuk menghilangkan rasa penat maupun mengisi waktu luang apalagi saat harus dirumah saja untuk mengurangi dampak penyebaran virus corona. Dari sini muncul masalah dari penikmat film yang kesulitan untuk menentukan film apa yang sesuai dengan kesukaannya entah itu berupa genre, opini maupun rating. Dari permasalahan

tersebut peneliti mencoba mengatasi menggunakan sistem rekomendasi.

Sistem rekomendasi adalah suatu sistem yang akan memberikan rekomendasi sesuai dengan apa yang sudah lakukan pengguna seperti pemilihan produk dengan tujuan dapat memberikan pilihan produk yang efektif sesuai dengan keinginan pengguna (Kurniawan, 2016). Kriteria yang digunakan dalam implementasi sistem rekomendasi bisa



berupa opini maupun rating yang diberikan oleh user (Jaja, Susanto, & Sasongko, 2020). Terdapat beberapa teknik sistem rekomendasi yaitu Collaborative Filtering (CF), Content Based Filtering (CBF), demografi dan sistem rekomendasi hybrid (Yadav, Vikesh, Shreyam, & Nagpal, 2018). Teknik Collaborative Filtering akan memprediksi rekomendasi berdasarkan hubungan antara pengguna, variabel rating dari item dan item yang disukai oleh pengguna (Shakirova, 2017). Sedangkan teknik Content Based Filtering merekomendasikan item serupa dengan kecenderungan pengguna yang tidak memiliki history item sebelumnya menggunakan data yang independent (Rizky, Asror, & Murti, 2020). Demografi sendiri adalah Teknik dari sistem rekomendasi yang berdasarkan data demografi dari user seperti pekerjaan, pendidikan, umur, jenis kelamin dan data demografis lainnya (Muhammad Alkaff, 2020). Sedangkan sistem rekomendasi hybrid adalah metode yang menggabungkan teknik – Teknik sistem rekomendasi menjadi satu (Tommy, Novianto, & Japriadi, 2020).

Dari teknik tersebut yang memberikan rekomendasi paling sesuai adalah teknik Collaborative Filtering selanjutnya dalam penelitian ini akan disebut menjadi CF (Yadav, Vikesh, Shreyam, & Nagpal, 2018). Dalam teknik CF terdapat dua jenis metode yaitu *User Based Collaborative Filtering* / UBCF dan *Content Based Collaborative Filtering* (CBCF). UBCF adalah metode yang memberikan rekomendasi berdasarkan kemiripan antar user-nya sedangkan CBCF berdasarkan parameter item yang disukai (Hadi, Santoso, & Tjondrowiguno, 2020). Peneliti sebelumnya mendapatkan hasil UBCF mempunyai akurasi yang lebih baik. Peneliti akan mencoba menggunakan UBCF dalam penelitian ini untuk mencari rekomendasi film sesuai dengan preferensi user.

Tahapan dalam UBCF adalah mencari nilai kemiripan antar user yang nantinya akan dijadikan bobot sebagai penyelesaian masalah sparsity. Metode yang sering digunakan dalam pencarian nilai kemiripan atau jarak antar user seperti mean-squared difference, Pearson correlation, cosine similarity, Spearman correlation, and adjusted cosine similarity (Fayyaz, Ebrahimian, Nawara, Ibrahim, & Kashef, 2020). Peneliti lain, mengatakan bahwa metode Pearson Correlation adalah metode yang tepat digunakan untuk menghitung jarak terdekat antara user berdasarkan rating. Peneliti sebelumnya juga menyimpulkan bahwa Pearson Correlation Coefficient menghasilkan peningkatan kualitas dalam menghitung dataset yang besar (Al-Bakri & Hashim, 2018).

Sehingga peneliti dalam penelitian ini menggunakan pearson untuk mencari nilai kemiripan.

Akan tetapi CF memiliki kekurangan ketika pengguna tidak memberikan rating atau tidak menyukai suatu item akan menyebabkan sparsity atau kelonggaran data user yang memiliki kemiripan yang berakibat menyulitkan proses pencarian rekomendasi (Yadav, Vikesh, Shreyam, & Nagpal, 2018). Peneliti sebelumnya, Sambhav Yadav menyimpulkan bahwa memberikan bobot per item pada matrik rating user dapat membantu menemukan kemiripan user dan membantu mengatasi sparsity. Penggunaan algoritma genetika dapat digunakan untuk membantu mencari kemiripan user. Algoritma genetika adalah algoritma dalam bidang optimasi pencarian yang efektif dalam ruang yang besar (Jepriana & Wardoyo, 2018).

Algoritma Firefly yaitu algoritma yang terinspirasi dari gerombolan kunang – kunang dan merupakan salah satu algoritma terpopuler untuk masalah optimasi (Hay's, 2017). Algoritma ini lebih unggul dalam menyelesaikan permasalahan optimasi dibandingkan dengan algoritma genetika (Arianti, Via, & Purbasari, 2020). Peneliti menggunakan algoritma ini untuk menentukan user – user mana yang mempunyai nilai kemiripan tinggi. Dalam penelitian ini menggunakan pearson correlation dan algoritma firefly dengan metode user based collaborative filtering untuk merekomendasi suatu film. Dengan mengkombinasikan algoritma firefly dan pearson correlation diharapkan dapat membantu menghasilkan rekomendasi yang lebih baik.

2. Metode Penelitian

Langkah pertama dimulai dari mempersiapkan dataset yaitu MovieLens. Selanjutnya adalah membuat matrik dari data rating – rating film yang diambil dari dataset. Matrik rating ini akan dihitung nilai kemiripan antar ratingnya dengan menggunakan metode *similarity* yaitu *pearson correlation coefficient*. Setelah itu sistem akan mengelompokkan user atau merangking user yang memiliki kemiripan tinggi menggunakan algoritma *firefly*. Lalu dari hasil rangking tersebut dilakukan prediksi rating dan yang terakhir adalah melakukan pengujian dari rating sebenarnya dan rating prediksi menggunakan *Mean Absolute Error* dan *Mean Squared Error*. Pada Gambar 1 menjelaskan alur dari metode penelitian yang dilakukan.

D. User Based Collaborative Filtering

Setelah semua metode sudah diimplementasikan, langkah selanjutnya adalah melakukan rekomendasi atau mencari film yang paling sesuai dengan user dengan teknik user based collaborative filtering. Berikut pada Gambar 6 contoh dari hasil rekomendasi dari user 414.

Film yang direkomendasikan untuk id user 414:

	title	movieId
291	Tommy Boy (1995)	291
277	Shawshank Redemption, The (1994)	277

Gambar 6. Hasil Rekomendasi

E. Pengujian

Langkah terakhir dari penelitian adalah melakukan pengujian dari hasil rekomendasi. Pengujian atau evaluasi hasil prediksi rating menggunakan metode *Mean Absolute Error* (MAE) dan *Mean Square Error* (MSE). MAE adalah metode untuk menghitung error berdasarkan nilai rata-rata kesalahan (Raghuwanshi & Pateriya, 2019). Sedangkan MSE adalah metode pencarian nilai error yang menggunakan nilai kuadrat dari kesalahan (Pranoto, Harianto, & Iswanto, 2020). Skenario evaluasi yang digunakan yaitu dengan jumlah data yang variatif serta user yang melakukan rating terbanyak dan user yang paling sedikit melakukan rating film. Data yang akan dievaluasi adalah hasil prediksi rating antara sistem rekomendasi menggunakan pearson dan yang menggunakan penggabungan pearson dan firefly supaya dapat mengetahui apakah ada peningkatan atau penurunan akurasi dari kedua metode tersebut.

Hasil evaluasi menggunakan metode MAE dapat dilihat pada Tabel 1. P menggunakan metode Pearson, PF menggunakan Pearson Firefly. Label user banyak adalah user yang banyak dalam melakukan rating dan user sedikit adalah user yang sedikit merating.

Tabel 1. Evaluasi MAE

Jumlah Data		1.000	10.000	100.000	
User Banyak	P	0.76	0.56	0.53	0.62
User Sedikit	P	0.76	0.56	0.53	0.62
					Rerata
					0.62

User Banyak	PF	0.95	0.8	0.73	0.83
User Sedikit	PF	0.95	0.8	0.73	0.83
					Rerata
					0.83

Dari Tabel 1 dapat disimpulkan bahwa evaluasi menggunakan metode MAE untuk user sedikit maupun user banyak menghasilkan akurasi yang sama dari jumlah data yang berbeda. Rata-rata akurasi MAE adalah sebesar 0.62 untuk metode pearson dan 0.83 untuk pearson firefly. Selanjutnya pada Tabel 2 menjabarkan hasil dari evaluasi menggunakan metode MSE dalam %.

Tabel 2. Evaluasi MSE

Jumlah Data		1.000	10.000	100.000	
User Banyak	P	80.11	85.27	86.19	83.86
User Sedikit	P	79.82	85.45	86.14	83.80
					Rerata
					83.83
User Banyak	PF	76.96	79.90	81.04	79.30
User Sedikit	PF	76.94	79.89	81.04	78.42
					Rerata
					78.86

Hasil evaluasi menggunakan MSE menghasilkan akurasi yang berbeda antar user. Rata-rata akurasinya adalah untuk pearson sebesar 83.83% dan untuk pearson firefly sebesar 78.86%

4. Kesimpulan

Dari hasil penelitian dengan skenario evaluasi menggunakan 1000 data, 10.000 data dan 100.000 data menghasilkan akurasi yang berbeda, semakin banyak data yang digunakan semakin baik pula akurasinya. Dengan menggunakan metode Pearson *Firefly* menghasilkan rerata MAE yang sama dengan jumlah user yang sedikit maupun yang banyak melakukan rating yaitu sebesar 0.83 dan rerata akurasi sebesar 78.86%. Penggunaan metode ini tidak menghasilkan optimasi yang tepat dalam penentuan kemiripan user. Metode pearson sendiri mendapatkan error rata-rata lebih rendah 0.21 yaitu sebesar 0.62 dan peningkatan rerata akurasi sebesar 83.83%. Dari penelitian ini dapat disimpulkan bahwa penggunaan algoritma *Firefly* tidak cocok untuk masalah pencarian kelompok dalam sistem rekomendasi. Sedangkan Pearson lebih baik dalam pencarian kelompok dan sistem

rekomendasi dengan hasil akurasi yang lebih tinggi. Peneliti lain dapat mencoba mengkolaborasikan pearson dengan metode atau algoritma lain untuk menemukan akurasi yang lebih baik.

Referensi

- Al-Bakri, N. F., & Hashim, S. H. (2018, June). Reducing Data Sparsity in Recommender Systems. *Journal of Al-Nahrain University*, 21 (2), 38-147.
- Arianti, R. W., Via, Y. V., & Purbasari, I. Y. (2020, Juli). Implementasi Algoritma Firefly Dalam Menyelesaikan Pengoptimalan Produksi Sepatu (Studi Kasus : Home industry "PAK KICU Shoes" Sidoarjo). *Jurnal Informatika dan Sistem Informasi (JIFoSI)*, 1.
- Fayyaz, Z., Ebrahimian, M., Nawara, D., Ibrahim, A., & Kashef, a. (2020). Recommendation Systems: Algorithms, Challenges, Metrics, and Business Opportunities. *Applied Science*.
- Hadi, I., Santoso, L. W., & Tjondrowiguno, A. N. (2020). Sistem Rekomendasi Film menggunakan User-based Collaborative Filtering dan K-modes Clustering. *JURNAL INFRA*.
- Hay's, R. N. (2017, Juni). Implementasi Firefly Algorithm-Tabu Search Untuk Penyelesaian Traveling Salesman Problem. *JOIN*, 2.
- Jaja, Y. V., Susanto, B., & Sasongko, L. R. (2020, September). Penerapan Metode Item-Based Collaborative Filtering Untuk Sistem Rekomendasi Data MovieLens. *d'Cartesian : Jurnal Matematika dan Aplikasi*, 9, 1-6.
- Jepriana, I. W., & Wardoyo, R. (2018). Algoritme Genetika untuk Mengurangi Galat Prediksi Metode Item-based Collaborative Filtering. *BIMIPA*, 25(2), 165-174.
- Kurniawan, A. (2016). Sistem Rekomendasi Produk Sepatu Dengan Menggunakan Metode Collaborative Filtering. *Seminar Nasional Teknologi Informasi Dan Komunikasi (SENTIKA 2016)*, 610-614.
- Muhammad Alkaff, H. K. (2020, November). Sistem Rekomendasi Buku Menggunakan Weighted Tree Similarity Dan Content Based Filtering. *Matrik : Jurnal Manajemen, Teknik Informatika Dan Rekayasa Komputer*, 20, 193-202.
- Pranoto, Y. M., Harianto, R. A., & Iswanto. (2020). Pemanfaatan Arima Untuk Prediksi Harga Emas Dalam Sistem Rekomendasi Trading Gold Option. *JURNAL MEDIA INFORMATIKA BUDIDARMA*, 863-871.
- Raghuwanshi, S. K., & Pateriya, R. K. (2019). Collaborative Filtering Techniques In Recommendation Systems. *Springer Nature Singapore*.
- Rizky, M. I., Asror, I., & Murti, Y. R. (2020, April). Sistem Rekomendasi Program Studi Untuk Siswa SMA Sederajat Menggunakan Metode Hybrid Recommendation Dengan Content Based Filtering Dan Collaborative Filtering. *E-Proceeding Of Engineering*, 7, 2276.
- Shakirova, E. (2017). Collaborative Filtering For Music Recommender System. *IEEE Conference Of Russian Young Researchers In Electrical And Electronic Engineering (Eiconrus)*.
- Tommy, L., Novianto, D., & Japriadi, Y. S. (2020, Desember). Sistem Rekomendasi Hybrid Untuk Pemesanan Hidanganberdasarkan Karakteristik Dan Rating Hidangan. *Journal Of Applied Informatics And Computing(JAIC)*, 4, 137-145.
- Yadav, S., Vikesh, Shreyam, & Nagpal, S. (2018). An Improved Collaborative Filtering Based Recommender System Using Bat Algorithm. *Procedia Computer Science*, 1795–1803.