

Algoritme Genetika untuk Mengurangi Galat Prediksi Metode Item-based Collaborative Filtering

*Genetic Algorithms for Reducing the Prediction Error
of Item-based Collaborative Filtering Methods*

I Wayan Jepriana ^{*1}, Retantyo Wardoyo ²

^{1,2}Jurusan Ilmu Komputer dan Elektronika, FMIPA UGM, Yogyakarta

e-mail: ^{*}jepriana@mail.ugm.ac.id, ²rw@ugm.ac.id

Abstrak

Metode yang banyak digunakan dalam sistem rekomendasi adalah *collaborative filtering*. Pengembangan dari pendekatan untuk mengurangi kesalahan prediksi sudah menjadi sebuah topik penelitian yang aktif dalam sistem rekomendasi dengan *collaborative filtering*. Akurasi dari prediksi memainkan peran yang sangat penting dalam pemberian rekomendasi kepada pengguna. Salah satu permasalahan yang terjadi pada *collaborative filtering* adalah jika hanya sedikit pengguna yang telah memberikan rating untuk *item*, yang menyebabkan sulit untuk mengidentifikasi *neighborhood* yang benar dan relevan untuk membuat prediksi. Terdapat dua macam metode *collaborative filtering* yaitu *user-based collaborative filtering* dan *item-based collaborative filtering*. Metode *item-based* memberikan kualitas prediksi yang lebih baik daripada metode *user-based*.

Optimasi metode *user-based collaborative filtering* dengan algoritme genetika untuk memperbaiki nilai kemiripan *user* mampu mengurangi galat prediksi. Pada penelitian ini dikembangkan optimasi sistem rekomendasi dengan algoritme genetika untuk memperbaiki nilai kemiripan *item*. Penelitian ini menggunakan *dataset movielens* dan *book-crossing*. Evaluasi dilakukan dengan variasi jumlah *neighbors* terhadap metode pengukur kemiripan yang digunakan, seperti: *cosine similarity*, *pearson correlation* dan *adjusted cosine*. Hasil menunjukkan bahwa usulan pendekatan mampu menurunkan galat prediksi pada seluruh metode pengukur kemiripan yang digunakan untuk kedua *dataset*. Galat prediksi terendah diperoleh evolusi metode *adjusted-cosine* pada kedua *dataset*.

Kata kunci: optimasi, sistem rekomendasi, *item-based collaborative filtering*, algoritme genetika

Abstract

A widely used method in recommender systems is collaborative filtering. The development of an approach for reducing the prediction error has been an active research field in collaborative filtering recommender systems. The accuracy of the predictions plays a very important role to provide recommendations to users. One of the problems with collaborative filtering methods is that only a few users have rated the items, which makes it difficult to identify the correct and relevant neighborhoods to make predictions. There are two types of collaborative filtering methods: user-based collaborative filtering and item-based collaborative filtering. The item-based method provides better prediction quality than the user-based method.

The optimization of user-based collaborative filtering method with genetic algorithm approach for refining similarity values can reduce prediction error. In this research, optimization of recommender system is developed by applying genetic algorithm for refining the similarity values between item. This research uses movielens and book-crossing datasets. Evaluation was done by variation number of neighbors to the similarity measuring method used, such as cosine similarity, pearson correlation and adjusted cosine. The results show that the proposed approach can reduce the prediction error in all the similarity measurement

methods used for both datasets. The lowest prediction error obtained by evolution of adjusted-cosine method on both datasets.

Keywords: optimization, recommender system, item-based collaborative filtering, genetic algorithm

1. Pendahuluan

Sistem rekomendasi adalah suatu teknik dan perangkat lunak yang dapat memberikan suatu anjuran atau rekomendasi yang sekiranya bermanfaat bagi pengguna dimana anjuran tersebut membantu dalam proses pengambilan keputusan oleh pengguna (Ricci dkk., 2011). Sedangkan menurut Naughton dan Lin (2010), sistem rekomendasi adalah sistem yang membantu pengguna dalam mengatasi informasi yang meluap dengan memberikan rekomendasi spesifik bagi pengguna dan diharapkan rekomendasi tersebut bisa memenuhi keinginan dan kebutuhan pengguna.

Collaborative Filtering adalah teknik sistem rekomendasi yang banyak digunakan dan mungkin merupakan yang paling umum. *Collaborative filtering* memanfaatkan informasi *rating* dari beberapa pengguna untuk memprediksi *rating item* untuk pengguna tertentu. Model preferensi biasanya terdiri dari vektor *item* yang diberikan *rating* oleh pengguna. Vektor dibandingkan dengan vektor pengguna lain dengan ukuran kemiripan yang sesuai dan *neighborhood* (lingkungan) dari pengguna yang serupa dapat diidentifikasi. Rekomendasi pada dasarnya terdiri dari *item* yang sebelumnya belum dilihat/belum diberikan *rating* dan terdapat pada *neighborhood* (Johansson, 2004). *Collaborative filtering* dibagi menjadi dua macam yaitu *user-based collaborative filtering* dan *item-based collaborative filtering* (Schafer dkk., 2007). Pada penelitian Sarwar dkk. (2001) dilakukan perbandingan terhadap dua pendekatan dalam *collaborative filtering*: algoritme *user-based* dengan algoritme *item-based*. Pada penelitian tersebut, disimpulkan bahwa algoritme *item-based* memberikan kinerja yang lebih baik daripada algoritme *user-based*, sementara pada saat yang sama algoritme *item-based* juga memberikan kualitas prediksi yang lebih baik dari pada algoritme *user-based*.

Pengembangan dari pendekatan untuk mengurangi kesalahan prediksi sudah menjadi sebuah topik penelitian yang aktif dalam sistem rekomendasi dengan *collaborative filtering*, karena akurasi dari prediksi memainkan peran yang sangat penting untuk preferensi kepada pengguna (Ar dan Bostanci, 2016). Salah satu permasalahan yang terjadi pada *collaborative filtering* adalah jika hanya ada sedikit pengguna yang telah memberikan *rating* untuk *item*, maka akan sulit untuk mengidentifikasi *neighborhood* yang benar dan relevan (Johansson, 2004). Algoritme genetika adalah metode optimasi pencarian yang efisien, adaptif dan kuat yang biasanya diterapkan di ruang-ruang pencarian yang sangat besar, kompleks dan multimodal (Bandyopadhyay dan Pal, 2007). Algoritme genetika telah banyak digunakan khususnya dalam dua aspek dari sistem rekomendasi: *clustering* dan *hybrid user models* (Bobadilla dkk., 2011).

Bobadilla dkk. (2011) mengusulkan metrik untuk mengukur nilai kemiripan antar *user* yang digunakan dalam sistem rekomendasi *collaborative filtering*. Usulan metrik diformulasikan melalui kombinasi linier sederhana antara nilai dan bobot. Nilai dihitung untuk setiap pasang *user* di antara setiap kemiripan diperoleh, sementara bobot hanya dihitung sekali, dengan menggunakan tahap sebelumnya dimana algoritme genetika mengekstrak pembobotan dari sistem rekomendasi yang bergantung pada sifat spesifik data dari masing-masing sistem rekomendasi. Hasil yang diperoleh menunjukkan adanya

peningkatan yang signifikan dalam kualitas prediksi, kualitas rekomendasi dan kinerja sistem.

Optimasi metode *collaborative filtering* dengan dengan algoritme genetika juga dilakukan oleh (Alhijawi dan Kilani, 2016). Algoritme genetika digunakan untuk menghitung nilai kemiripan antara masing-masing *user* tanpa menggunakan algoritme *similarity functions* yang sudah dikenal seperti *Pearson correlation* dan *vector cosine-based similarity*. Hasil yang diperoleh dari penelitian ini adalah 46% peningkatan kualitas prediksi dan 38% peningkatan performa dari sistem rekomendasi jika dibandingkan dengan teknik yang lainnya.

Algoritme genetika juga digunakan untuk mengatasi perbedaan variabel input dari proses penggabungan teknik yang berbeda oleh Fong dkk. (2008). Dalam tulisan ini dicetuskan pendekatan berbasis GA untuk mendukung model *hybrid* antara *collaborative filtering* dan *content-based filtering*. Secara khusus, ditunjukkan bahwa bagaimana variabel input dapat dikodekan ke dalam kromosom GA dalam berbagai modus. Untuk pengujian terhadap rancangan sistem dilakukan eksperimen dengan input data yang diambil dari MovieLens dan IMDB.

Pada penelitian Ar dan Bostanci (2016) diusulkan metode *hybrid* algoritme genetika dan *collaborative filtering*. Tidak seperti metode *user-based* konvensional yang secara langsung menggunakan nilai kemiripan antar *user* hasil dari proses perhitungan, digunakan pendekatan algoritme genetika untuk memperbaiki nilai kemiripan sebelum digunakan dalam proses prediksi. Hasil penelitian tersebut menunjukkan bahwa pendekatan evolusioner telah secara signifikan mengurangi kesalahan prediksi dengan menggunakan bobot berevolusi.

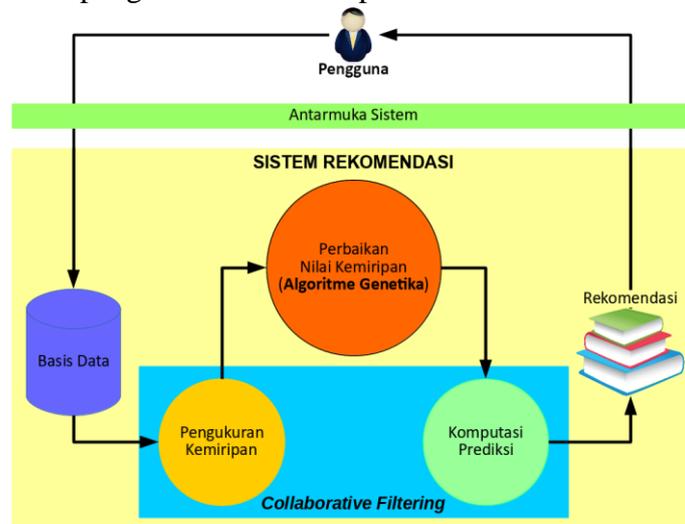
Berdasarkan permasalahan yang ada, pada penelitian ini dilakukan pengembangan metode *hybrid* untuk sistem rekomendasi pada metode *item-based collaborative filtering* dengan menerapkan algoritme genetika sebagai metode untuk memperbaiki nilai kemiripan antar *item* sebelum digunakan dalam proses komputasi prediksi.

2. Metode Penelitian

2.1 Rancangan Sistem

Rancangan sistem ditunjukkan pada Gambar 1. Untuk memperoleh rekomendasi, diperlukan input berupa data pengguna dan data *rating* yang ada pada basis data. Berdasarkan data *rating* yang telah diberikan pengguna kepada sejumlah *item*, akan dilakukan perhitungan nilai kemiripan antar *item* dengan menggunakan metode pengukur kemiripan seperti *cosine similarity*, *pearson correlation* atau *adjusted cosine similarity*. Perbaikan terhadap nilai kemiripan hasil perhitungan dilakukan perbaikan dengan algoritme genetika yang bertujuan untuk menurunkan galat prediksi *rating*. Untuk mendapatkan hasil prediksi akhir dibutuhkan suatu perhitungan yang dapat menghubungkan antara tingkat kemiripan *item* dengan *rating* yang telah diberikan oleh pengguna, pada penelitian ini digunakan metode *weighted sum*. Keluaran dari sistem adalah berupa prediksi *rating* yang dijadikan sebagai rekomendasi *item* untuk dapat menjadi bahan pertimbangan bagi pengguna. Keluaran sistem yang berupa prediksi *rating* akan digunakan sebagai bahan pengujian. Keluaran sistem akan dievaluasi dengan menghitung nilai *mean absolute error* (MAE). Pada perhitungan MAE prediksi *rating* dibandingkan dengan *rating* sebenarnya yang telah diberikan oleh *user*. Hasil evaluasi berupa nilai MAE dari prediksi *rating* terhadap *rating* sebenarnya yang telah diberikan oleh *user*. Hasil evaluasi akan dibandingkan

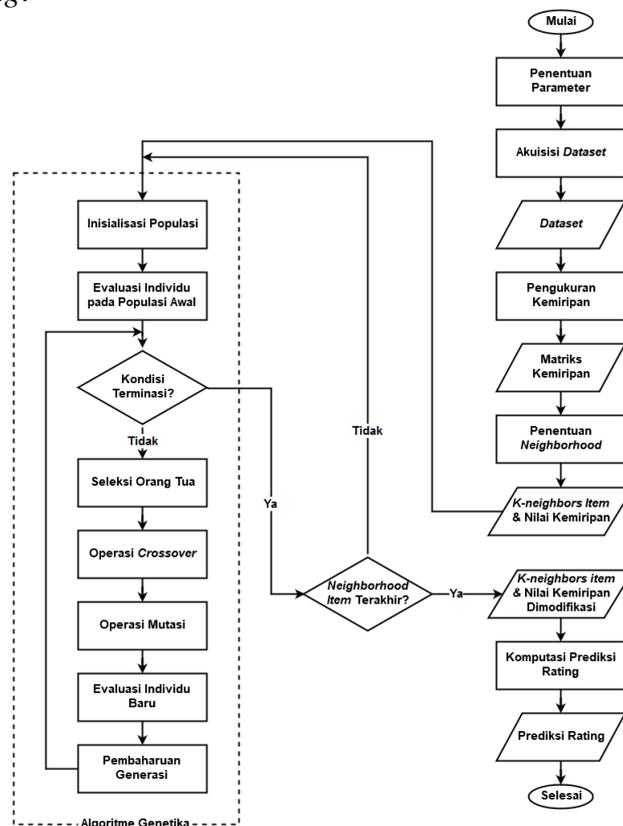
dengan galat prediksi nilai kemiripan tanpa proses evolusi dengan algoritme genetika untuk mengetahui pendekatan pengukur nilai kemiripan terbaik.



Gambar 1. Rancangan sistem

2.2 Rancangan Proses

Gambar 2 merupakan alur dari sistem rekomendasi yang dibagi menjadi dua bagian utama, yaitu bagian *collaborative filtering* yang menggunakan pendekatan *item-based collaborative filtering* dan bagian algoritme genetika yang merupakan proses perbaikan nilai kemiripan *item* yang akan digunakan pada proses komputasi prediksi *item-based collaborative filtering*.



Gambar 2. Alur Sistem

Alur sistem diawali dengan menentukan parameter yang merupakan parameter-parameter pada algoritme genetika. Kemudian dilanjutkan dengan akuisisi *dataset* yang akan digunakan. Berdasarkan *dataset* terpilih dilakukan pengukuran kemiripan antar *item* yang akan menghasilkan matriks kemiripan *item*. Matriks kemiripan akan digunakan sebagai landasan untuk perbaikan nilai kemiripan dengan algoritme genetika. Hasil dari proses perbaikan nilai kemiripan adalah berupa *k-neighbors* beserta nilai kemiripan yang sudah dimodifikasi untuk masing-masing *item* pada *dataset*. Keluaran sistem akan digunakan untuk melakukan komputasi prediksi *rating* yang akan dijadikan sebagai landasan rekomendasi untuk pengguna.

2.3 Desain Algoritma Genetika

Berdasarkan analisa terhadap rumusan masalah, bagian ini merupakan fokus pada penelitian ini. Algoritme genetika digunakan untuk memperbaiki nilai kemiripan antar *item* yang akan digunakan untuk menghasilkan rekomendasi pada metode *item-based collaborative filtering*. Berikut penjelasan untuk masing-masing komponen algoritme genetika yang digunakan pada penelitian ini.

Representasi kromosom

Kromosom direpresentasikan ke dalam bentuk *array* dengan panjang yang ditentukan berdasarkan jumlah *k-neighbors* maksimal yang akan digunakan. Kromosom dalam penelitian ini merupakan representasi dari nilai kemiripan suatu *item* dengan sejumlah *k* *item* lainnya yang ada pada *dataset*. Jika *j* merupakan sejumlah *item* yang merupakan *neighborhood* dari *item i*, dan *neighborhood* dari *item i* adalah sejumlah *k* maka suatu kromosom terdiri dari baris $sim(i, j_k)$. $sim(i, j_k)$ merupakan nilai kemiripan antara *item i* *item j_k* yang merupakan *neighborhood* dari *item i*. Kromosom dikodekan dengan menggunakan skema pengkodean nilai, dimana nilai yang digunakan dalam penelitian ini adalah rentang nilai [0...1] yang merupakan representasi dari nilai $sim(i, j_k)$. Gambar 3 merupakan ilustrasi representasi dari sebuah kromosom individu untuk *item i*. Kumpulan individu akan membentuk sebuah populasi.

Sim(i,j1)	Sim(i,j2)	Sim(i,j3)	Sim(i,j4)	...	Sim(i,jk)
-----------	-----------	-----------	-----------	-----	-----------

Gambar 3. Representasi kromosom

Inisialisasi populasi

Inisialisasi populasi akan dilakukan dengan menggunakan nilai kemiripan yang dihasilkan dari proses pengukuran nilai kemiripan dengan metode yang akan dievolusi. Sebagai tambahan anggota populasi akan dibangkitkan nilai secara acak untuk setiap gen pada masing-masing kromosom. Ukuran populasi ditetapkan 50.

Fungsi fitness

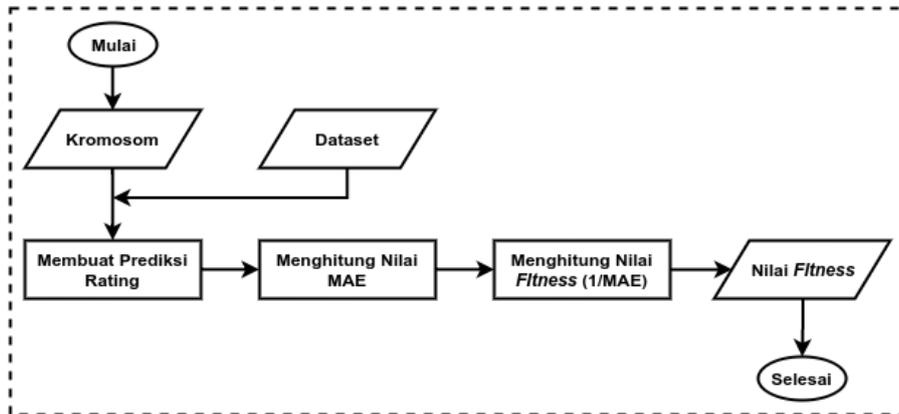
Proses evaluasi dilakukan untuk mencari individu terbaik dalam populasi. Individu terbaik didapatkan dengan menghitung nilai *fitness* masing-masing individu. Nilai *fitness* dihitung menggunakan fungsi *fitness* yang bertujuan untuk mengukur seberapa optimal masing-masing individu yang berada dalam populasi. Dalam rangka untuk menghitung nilai *fitness* untuk setiap individu, terlebih dahulu ditentukan fungsi objektif dari permasalahan. Fungsi objektif pada penelitian ini adalah meminimalkan nilai *error* dari prediksi *rating* dihasilkan oleh setiap individu terhadap *rating* yang sebenarnya terdapat pada data latih.

$$MAE = \frac{\sum_{i=1}^N |p_i - q_i|}{N} \quad (1)$$

Pada penelitian ini *error* yang digunakan adalah MAE seperti yang dirumuskan pada Persamaan (1). N adalah jumlah pengguna yang telah memberikan *rating* terhadap *item* yang sedang dievaluasi, p adalah prediksi yang dihasilkan oleh kromosom individu dan q adalah nilai *rating* sebenarnya yang berada pada data latih. Berdasarkan nilai MAE, agar individu yang memiliki nilai MAE paling kecil dapat memiliki nilai *fitness* paling besar, fungsi *fitness* dihitung dengan menggunakan Persamaan (2).

$$Fitness\ function = \frac{1}{MAE} \quad (2)$$

Gambar 4 merupakan alur proses perhitungan nilai *fitness* suatu kromosom individu.



Gambar 4. Alur proses menghitung nilai fitness

Operasi genetik

Berdasarkan nilai *fitness* untuk masing-masing individu, dilakukan seleksi individu yang akan dijadikan sebagai calon orang tua. Individu-individu yang terpilih pada proses seleksi orang tua akan disimpan ke populasi baru tentatif atau yang biasa disebut dengan *mating pool*. Seleksi pada penelitian ini akan menggunakan metode *roulette-wheel* untuk memilih individu yang kemungkinan akan digunakan pada proses *crossover*. Berdasarkan individu-individu hasil proses seleksi yang berada pada *mating pool*, akan dilakukan proses *crossover*. Jenis *crossover* yang digunakan pada penelitian ini adalah *whole arithmetic recombination* seperti yang dirumuskan pada Persamaan (3). Pendekatan ini digunakan karena dapat menciptakan material gen baru secara menyeluruh berdasarkan kromosom induk. Nilai α yang digunakan adalah 0,4.

$$\begin{aligned} Anak\ 1 &= \alpha \cdot \bar{x} + (1 - \alpha) \cdot \bar{y} \\ Anak\ 2 &= \alpha \cdot \bar{y} + (1 - \alpha) \cdot \bar{x} \end{aligned} \quad (3)$$

Operasi mutasi diterapkan pada anak-anak hasil dari operasi *crossover* dengan probabilitas mutasi 0,01. Jenis mutasi yang akan digunakan adalah mutasi *uniform*. Mutasi *uniform* dipilih karena merupakan salah satu operator mutasi untuk pengkodean bilangan riil yang dapat mendistribusikan solusi secara luas ke seluruh ruang pencarian.

Pembaharuan generasi dan kondisi terminasi

Individu baru hasil dari operasi *crossover* dan mutasi akan menggantikan anggota terburuk dalam populasi ketika mereka memiliki nilai *fitness* yang lebih baik. Jika individu baru tidak memiliki nilai *fitness* yang lebih baik dari anggota terburuk dalam populasi, maka akan dibuang. Untuk mengevaluasi algoritme genetika diperlukan kondisi dimana evolusi harus dihentikan, pada penelitian ini akan digunakan kondisi dimana jumlah generasi mencapai jumlah yang akan ditentukan berdasarkan pengamatan parameter optimal untuk *dataset*.

2.4 Data Penelitian

Data yang akan digunakan untuk pengujian metode pada penelitian ini adalah *dataset* rekomendasi film *MovieLens* dan *dataset* rekomendasi buku *Book-Crossing*. Data input pada penelitian ini adalah berupa file CSV dari masing-masing *dataset*.

Dataset movielens

Dataset movieLens banyak digunakan di bidang pendidikan, penelitian, dan industri. *Dataset movieLens* telah diunduh ratusan ribu kali setiap tahun, yang mencerminkan penggunaannya dalam buku-buku program pers populer, kursus tradisional dan online, dan perangkat lunak. *Dataset* ini adalah produk aktivitas anggota dalam sistem rekomendasi film *MovieLens*, sebuah *platform* penelitian aktif yang telah menyelenggarakan banyak percobaan sejak diluncurkan pada tahun 1997 (Harper dan Konstan, 2016). *Dataset* ini berisi 100,004 *rating* dan 1.296 aplikasi *tag* di 9.125 film. Data ini dibuat oleh 671 pengguna antara tanggal 09 Januari 1995 sampai 16 Oktober 2016. Sebagai catatan, *dataset* yang digunakan adalah *dataset* versi *small*. Pada penelitian ini, jumlah data akan dikurangi untuk mempercepat proses penelitian. Data yang akan digunakan adalah 625 *user*, 1.283 film dan 68.017 *rating* yang diberikan terhadap sejumlah film. Untuk setiap *user* telah memberikan minimal 20 *rating* terhadap *item*, dan setiap *item* setidaknya telah diberikan *rating* oleh 20 *user*.

Dataset book-crossing

Book-Crossing merupakan *dataset* yang dikumpulkan oleh Cai-Nicolas Ziegler (Agustus/September 2004) dari komunitas *Book-Crossing* dengan izin dari Ron Hornbaker yang merupakan *chief technology officer* (CTO) dari *Humankind Systems*. *Dataset book-crossing* terdiri dari 278.858 pengguna (anonim tetapi dengan informasi demografis) menyediakan 1.149.780 *rating* (eksplisit/implisit) tentang 271.379 buku. Pada penelitian Ziegler dkk. (2005) yang berjudul *Improving Recommendation Lists Through Topic Diversification*, *dataset book-crossing* digunakan. Pada penelitian ini, jumlah data akan dikurangi untuk mempercepat proses penelitian. Data yang akan digunakan adalah 1.037 *user* dan 811 *item* dengan banyak *rating* 20.639. Untuk setiap *user* telah memberikan minimal 10 *rating* terhadap *item* dan setiap *item* setidaknya telah diberi *rating* oleh 10 *user*.

3. Hasil dan Pembahasan

Sebelum melakukan pengujian dilakukan penentuan panjang kromosom yang digunakan pada penelitian ini, hasil pengujian menunjukkan jika panjang kromosom optimal untuk *dataset movielens* adalah 500, dan 800 untuk *dataset book-crossing*. Tabel 1 merupakan hasil pengujian dari masing-masing metode pengukur kemiripan pada *dataset movielens*. Dilakukan pengujian pada masing-masing metode pengukur kemiripan, baik bobot asli maupun berevolusi sebanyak 5 kali. Pada metode pengukur kemiripan dengan

bobot asli *adjusted cosine* menghasilkan rata-rata galat prediksi terendah yang ditandai dengan nilai MAE yang paling kecil.

Pada Tabel 1 juga menampilkan rata-rata galat prediksi yang dihasilkan masing-masing metode pengukur kemiripan setelah dilakukan proses evolusi dengan algoritme genetika. Rata-rata galat prediksi oleh metode *adjusted cosine* menghasilkan galat prediksi yang paling rendah secara keseluruhan. Jika dibandingkan dengan galat prediksi masing-masing metode tanpa proses evolusi, semua metode pengukur kemiripan mengalami penurunan galat prediksi dengan dilakukannya proses evolusi pada nilai kemiripan sebelum digunakan pada proses prediksi.

Tabel 1. Rata-rata nilai MAE pada dataset *movielens*

K-Neigh.	Cosine Similarity		Pearson Correlation		Adjusted Cosine	
	Asli	Berevolusi	Asli	Berevolusi	Asli	Berevolusi
20	0,7407	0,7170	0,7474	0,7115	0,6917	0,6809
40	0,7269	0,7166	0,7320	0,7122	0,6797	0,6759
60	0,7223	0,7174	0,7263	0,7137	0,6762	0,6747
80	0,7206	0,7179	0,7233	0,7146	0,6752	0,6744
100	0,7199	0,7182	0,7215	0,7154	0,6746	0,6740
120	0,7198	0,7184	0,7204	0,7158	0,6745	0,6739
140	0,7201	0,7187	0,7194	0,7162	0,6745	0,6738
160	0,7204	0,7188	0,7189	0,7163	0,6749	0,6737
180	0,7206	0,7188	0,7188	0,7166	0,6750	0,6737
200	0,7208	0,7188	0,7187	0,7166	0,6754	0,6737

Gambar 5 menampilkan hasil evaluasi dari 5 kali pengujian untuk masing-masing metode pengukur kemiripan dengan bobot asli dan berevolusi untuk *dataset movielens*. Semua evolusi dari nilai kemiripan menghasilkan penurunan galat prediksi yang signifikan. Hal itu ditandai dengan nilai *t* yang dihasilkan masing-masing perbandingan lebih besar dari pada t_{tabel} untuk $df=49$ dan $\alpha=0,05$ yaitu $t_{(0,05;49)} = 2,009575$. Selain itu juga nilai Sig. masing-masing perbandingan juga mendekati 0 yang menandakan bahwa keyakinan terhadap pengujian lebih besar dari 99% secara statistik.

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower	Upper			
Pair 1	Cosine Similarity Asli - Cosine Similarity Evo	.0051855	.0068130	.0009635	.0032493	.0071218	5.382	49	.000
Pair 2	Pearson Correlation Asli - Pearson Correlation Evo	.0097754	.0111662	.0015791	.0066020	.0129488	6.190	49	.000
Pair 3	Adjusted Cosine Asli - Adjusted Cosine Evo	.0022852	.0030920	.0004373	.0014064	.0031639	5.226	49	.000

Gambar 5. Hasil pengujian t-test dataset *movielens*

Pada Tabel 2 merupakan hasil pengujian dari masing-masing metode pengukur kemiripan pada *dataset book-crossing*. Dilakukan pengujian pada masing-masing metode pengukur kemiripan, baik bobot asli maupun berevolusi sebanyak 5 kali. Pada metode pengukur kemiripan dengan bobot asli *adjusted cosine* menghasilkan rata-rata galat prediksi

terendah pada jumlah *neighbors* 5 dan 10. Ketika jumlah *neighbors* bertambah sampai dengan 50 metode *pearson correlation* menghasilkan galat prediksi terendah. Kemudian setelah dilakukan proses evolusi dengan algoritme genetika, rata-rata galat prediksi oleh metode *adjusted cosine* menghasilkan galat prediksi yang paling rendah secara keseluruhan. Jika dibandingkan dengan galat prediksi masing-masing metode tanpa proses evolusi, semua metode pengukur kemiripan mengalami penurunan galat prediksi dengan dilakukannya proses evolusi pada nilai kemiripan sebelum digunakan pada proses prediksi.

Tabel 2. Rata-rata nilai MAE pada *dataset book-crossing*

K-Neigh.	Cosine Similarity		Pearson Correlation		Adjusted Cosine	
	Asli	Berevolusi	Asli	Berevolusi	Asli	Berevolusi
5	1,3115	1,2371	1,2355	1,2303	1,2324	1,2232
10	1,2791	1,1994	1,2035	1,1959	1,2018	1,1926
15	1,2697	1,1922	1,1967	1,1902	1,1975	1,1885
20	1,2676	1,1900	1,1949	1,1893	1,1967	1,1867
25	1,2660	1,1887	1,1943	1,1883	1,1967	1,1859
30	1,2656	1,1885	1,1939	1,1881	1,1963	1,1857
35	1,2646	1,1881	1,1938	1,1879	1,1963	1,1854
40	1,2648	1,1879	1,1938	1,1879	1,1965	1,1852
45	1,2645	1,1875	1,1936	1,1879	1,1965	1,1852
50	1,2645	1,1873	1,1936	1,1879	1,1965	1,1852

Gambar 6 menampilkan hasil evaluasi dari 5 kali pengujian untuk masing-masing metode pengukur kemiripan dengan bobot asli dan berevolusi pada *dataset book-crossing*. Semua evolusi dari nilai kemiripan menghasilkan penurunan galat prediksi yang signifikan. Hal itu ditandai dengan nilai t yang dihasilkan masing-masing perbandingan lebih besar dari pada t_{tabel} untuk $df=49$ dan $\alpha=0,05$ yaitu $t_{(0,05;49)} = 2,009575$. Selain itu juga nilai Sig. masing-masing perbandingan juga mendekati 0 yang menandakan bahwa keyakinan terhadap pengujian lebih besar dari 99% secara statistik.

Paired Samples Test

		Paired Differences				t	df	Sig. (2-tailed)	
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference				
					Lower				Upper
Pair 1	Cosine Similarity Asli - Cosine Similarity Evo	.0771289	.0065658	.0009285	.0752629	.0789949	83.064	49	.000
Pair 2	Pearson Correlation Asli - Pearson Correlation Evo	.0059961	.0046855	.0006626	.0046645	.0073277	9.049	49	.000
Pair 3	Adjusted Cosine Asli - Adjusted Cosine Evo	.0103516	.0038257	.0005410	.0092643	.0114388	19.133	49	.000

Gambar 6. Hasil pengujian t -test *dataset book-crossing*

4. Kesimpulan

Penelitian ini menunjukkan bahwa perbaikan nilai kemiripan dengan algoritme genetika yang dilakukan pada nilai kemiripan *item* hasil dari pengukuran dengan metode *cosine similarity*, *pearson correlation* dan *adjusted cosine similarity* dapat menurunkan galat prediksi pada sistem rekomendasi *item-based collaborative filtering* berdasarkan nilai *mean absolute error*. Nilai kemiripan yang berevolusi dapat memberikan penurunan galat prediksi

yang signifikan berdasarkan statistik. Evolusi nilai kemiripan yang dilakukan pada metode *adjusted cosine* menghasilkan galat prediksi terendah baik pada kedua *dataset* yang digunakan.

Daftar Pustaka

- Alhijawi, B., dan Kilani, Y., 2016, Using genetic algorithms for measuring the similarity values between users in collaborative filtering recommender systems. In *Computer and Information Science (ICIS), 2016 IEEE/ACIS 15th International Conference on* (pp. 1-6). IEEE.
- Ar, Y., & Bostanci, E., 2016, A genetic algorithm solution to the collaborative filtering problem. *Expert Systems with Applications*, 61, 122-128.
- Bandyopadhyay, S., dan Pal, S. K., 2007, *Classification and learning using genetic algorithms: applications in bioinformatics and web intelligence*, Springer Science & Business Media.
- Bobadilla, J., Ortega, F., Hernando, A., & Alcalá, J., 2011, Improving collaborative filtering recommender system results and performance using genetic algorithms, *Knowledge-based systems*, 24(8), 1310-1316.
- Fong, S., Ho, Y., dan Hang, Y., 2008, Using genetic algorithm for hybrid modes of collaborative filtering in online recommenders, *2008 Eighth International Conference on Hybrid Intelligent Systems*, (pp. 174-179), IEEE.
- Harper, F. M., dan Konstan, J. A., 2016, The movielens datasets: History and context, *ACM Transactions on Interactive Intelligent Systems (TiiS)*, 5(4), 19.
- Johansson, P., 2004, Design and Development of Recommender Dialogue Systems, *Desertasi*, Institutionen för datavetenskap.
- Naughton, R. dan Lin, X., 2010, Recommender Systems: Investigating the Impact of Recommendations on User Choices and Behaviors, *Proceedings of the ACM RecSys Workshop on User-Centric Evaluation of Recommender Systems and Their Interfaces*, Barcelona, Spain.
- Ricci, F., Rokach, L., dan Shapira, B., 2011, Introduction to Recommender Systems, *Recommender Systems Handbook*, Springer, New York, 1-35.
- Sarwar, B., Karypis, G., Konstan, J., dan Riedl, J., 2001, Item-Based Collaborative Filtering Recommendations Algorithm, *Proceeding of 10th International World Wide Web Conference*, ACM Press, 285-295.
- Schafer, J. B., Frankowski, D., Herlocker, J., dan Sen, S., 2007, Collaborative filtering recommender systems, *The adaptive web* (pp. 291-324), Springer Berlin Heidelberg.
- Ziegler, C. N., McNee, S. M., Konstan, J. A., dan Lausen, G., 2005, Improving recommendation lists through topic diversification, *Proceedings of the 14th international conference on World Wide Web* (pp. 22-32), ACM Chiba, Japan 10-14 2005.