

“PENERAPAN NON-NEGATIVE MATRIX FACTORIZATION DAN FUZZY C-MEANS DALAM MENGATASI DATA SPARSITY PADA SISTEM REKOMENDASI”

Moh. Nurun Fitriantama
(150411100082)

Dosen Pembimbing 1 : Dr. Noor Ifada, S.T., M.ISD.
Dosen Pembimbing 2 : Moch. Kautsar Sophan, S.Kom., M.MT.

ABSTRAK

Dengan berkembangnya teknologi yang pesat dan informasi yang berukuran besar, tentunya pengguna mengalami kesulitan ketika akan mencari informasi. Maka dari itu dibutuhkan sebuah sistem rekomendasi yang dapat memberikan informasi yang relevan dengan pengguna. *Collaborative Filtering* (CF) adalah salah satu pendekatan yang sering digunakan dalam sistem rekomendasi. CF biasanya digunakan untuk mengatasi permasalahan *sparsity*. *Sparsity* adalah keadaan dimana terdapat banyaknya data atau nilai rating yang kosong. Pendekatan CF memiliki tiga kategori yaitu *memory-based*, *model-based*, dan *hybrid*. Tentunya dari beberapa kategori itu memiliki beberapa kelemahan, seperti *memory-based* yang memberikan akurasi yang kurang optimal ketika menggunakan dataset *sparsity*. Tujuan penelitian ini adalah mengatasi permasalahan *sparsity* menggunakan CF yang menerapkan *Non-Negative Matrix Factorization* (NMF) dan *Fuzzy C-Means Clustering* (FCM). NMF digunakan sebagai pengurangan dimensi dengan nilai optimal dimensi (optimal 'k') dan mempertahankan nilai positif. FCM digunakan sebagai pengelompokan antara pengguna aktif dengan pengguna lainnya yang berada dalam satu kelompok. Pada penelitian ini, akan menggunakan metrik evaluasi *Normalized Discounted Cumulative Gain* (NDCG) yang digunakan untuk mengetahui hasil evaluasi dari pendekatan yang diusulkan. Berdasarkan skenario pengujian yang dilakukan untuk mengatasi masalah data *sparsity*, menghasilkan nilai $NDCG@10=0,0755$ atau 7,55%. Pendekatan CF yang menerapkan NMF dan FCM menunjukkan hasil yang lebih baik dibandingkan metode *Item-based Collaborative Filtering* (IBCF) dengan rata-rata kenaikan akurasi $NDCG=31,41\%$, maksimal $NDCG=40,17\%$ dan minimal $NDCG=20,28\%$.

Kata kunci : Sistem Rekomendasi, *Collaborative Filtering*, *Sparsity*, *Non-Negative Matrix Factorization* (NMF), *Fuzzy C-Means Clustering* (FCM), *Item-based Collaborative Filtering* (IBCF), *Normalized Discounted Cumulative Gain* (NDCG).

KATA PENGANTAR

Alhamdulillah, puji syukur penulis panjatkan kepada Allah SWT atas berkat dan rahmat-Nya penulis dapat menyelesaikan skripsi yang berjudul “Penerapan *Non-Negative Matrix Factorization* dan *Fuzzy C-Means* dalam Mengatasi Data *Sparsity* pada Sistem Rekomendasi”. Dengan terselesaikannya laporan skripsi ini, penulis tak lupa menyampaikan terima kasih yang sebesar-besarnya kepada :

1. Allah SWT yang telah melimpahkan rahmat dan keberkahan bagi penulis sehingga penulis dapat merasakan nikmat yang diberikan sampai saat ini dan Nabi Muhammad SAW yang telah menuntun umat manusia dari kebodohan menuju cahaya Islam.
2. Keluarga yang selalu memberikan doa, semangat, kasih sayang dan motivasi baik moral maupun materi.
3. Ibu Dr. Noor Ifada, S.T., M.ISD. dan Bapak Moch. Kautsar Sophan, S.Kom., M.MT., selaku dosen pembimbing yang selalu memberikan arahan dan bimbingan selama penelitian skripsi ini.
4. Segenap staf dan dosen di Teknik Informatika yang telah membimbing penulis selama masa perkuliahan.

Penulis menyadari bahwa pengerjaan skripsi ini tidak luput dari berbagai kesalahan. Oleh karena itu, penulis menyatakan permohonan maaf atas segala kekurangan yang ada. Dengan kerendahan hati, berbagai saran dan kritik yang membangun sangat penulis harapkan dari rekan-rekan pembaca.

Bangkalan, Januari 2020

Penulis

DAFTAR ISI

HALAMAN JUDUL.....	i
HALAMAN PENGESAHAN.....	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
ABSTRAK	iv
KATA PENGANTAR	v
DAFTAR ISI.....	vi
DAFTAR GAMBAR	ix
DAFTAR TABEL	x
DAFTAR KODE PROGRAM.....	xi
BAB I PENDAHULUAN	1
1.1. Latar Belakang	1
1.2. Perumusan Masalah.....	3
1.2.1 Permasalahan.....	3
1.2.2 Metode Usulan	3
1.2.3 Pertanyaan Penelitian	3
1.3. Tujuan dan Manfaat.....	3
1.3.1 Tujuan Penelitian	3
1.3.2 Manfaat	4
1.4. Batasan masalah	4
1.5. Sistematika Proposal	4
BAB II KAJIAN PUSTAKA	5
2.1. Sistem Rekomendasi	5
2.1.1 Content-Based	5
2.1.2 Collaborative Filtering	7
2.1.3 Hybrid	8
2.2. Non-Negative Matrix Factorization	10
2.3. Fuzzy C-Means Clustering.....	11
2.4. Pearson Coefficient Corelation (PCC).....	13
2.5. Pendekatan Evaluasi	14
2.6. Metrik Evaluasi	14

2.7. Penelitian Terkait	15
BAB III METODE USULAN.....	17
3.1. Arsitektur Rancangan Sistem	17
3.2. Dataset	18
3.3. Skenario Pengujian.....	19
3.4. Alur Kinerja Pengujian Sistem.....	20
3.5. Contoh Perhitungan	21
3.5.1 Input Data Rating Menjadi Matriks $R \in \mathbb{R}^{10 \times 6}$	21
3.5.2 Proses NMF pada Matriks $R \in \mathbb{R}^{10 \times 6}$	21
3.5.3 Proses Pengelompokan dengan FCM	22
3.5.4 Hitung Kemiripan Target <i>User</i> dengan <i>User</i> Lainnya yang Berada Dalam Satu <i>Cluster</i>	28
3.5.5 Tentukan Top- <i>m</i> Neighbour, Hitung Prediksi Rating, dan Tentukan top- <i>n</i> rekomendasi.....	29
BAB IV IMPLEMENTASI METODE.....	30
4.1. Lingkungan Uji Coba	30
4.2. Dataset	30
4.3. Tahapan Pembuatan Program.....	31
4.3.1 Tahapan berdasarkan IPO (<i>Input – Proses – Output</i>)	31
4.3.2 Proses Evaluasi Matriks	40
4.4. Skenario dan Hasil Uji Coba	41
4.4.1 Skenario Uji Coba Variasi Rank (<i>f</i>)	41
4.4.2 Skenario Uji Coba Variasi Cluster (<i>c</i>)	42
4.4.3 Skenario Uji Coba Variasi Top- <i>m</i> Neighbour.....	42
4.5. Perbandingan Pendekatan.....	43
4.5.1 Pendekatan Pembandingan (IBCF).....	44
4.5.2 Perbandingan Pendekatan dengan Matriks Evaluasi NDCG	44
4.6. Tampilan <i>User Interface</i> Sistem.....	45
BAB V PENUTUP.....	50
5.1. Kesimpulan.....	50
5.2. Saran	50
REFERENSI	51

DAFTAR RIWAYAT HIDUP.....	55
---------------------------	----

DAFTAR GAMBAR

Gambar 2.1 Gambaran Umum Sistem Rekomendasi	5
Gambar 2.2 Tampilan pandora radio	6
Gambar 2.3 Ilustrasi dari Pendekatan Content-Based.....	6
Gambar 2.4 Ilustrasi dari pendekatan Collaborative Filtering	7
Gambar 2.5 Tampilan rekomendasi teman pada facebook	7
Gambar 2.6 Tampilan rekomendasi video pada youtube.....	9
Gambar 2.7 Ilustrasi Proses NMF	10
Gambar 2.8 Ilustrasi pendekatan n-fold cross validation.....	14
Gambar 3.1 Diagram IPO	18
Gambar 3.2 Skema Pengujian Sistem	20
Gambar 4.1 Grafik perbandingan pendekatan berdasarkan matriks evaluasi NDCG	45
Gambar 4.2 Tampilan User Interface Sistem (NFCF)	46
Gambar 4.3 Tampilan User Interface Sistem (IBCF)	48

DAFTAR TABEL

Tabel 2.1 Perbandingan Pendekatan Pada Sistem Rekomendasi.....	9
Tabel 2.2 Penelitian sebelumnya yang menggunakan teknik <i>Dimensionality Reduction</i> dan <i>Clustering</i>	16
Tabel 3.1 Contoh dataset MovieLens	19
Tabel 3.2 Skenario Pengujian	19
Tabel 3.3 Matriks $R \in \mathbb{R}^{10 \times 6}$	21
Tabel 3.4 Matriks Latent $U \in \mathbb{R}^{10 \times 3}$	21
Tabel 3.5 Matriks Latent $I \in \mathbb{R}^{3 \times 6}$	22
Tabel 3.6 Matriks partisi awal iterasi 1	22
Tabel 3.7 Perhitungan pusat cluster 1 iterasi 1	23
Tabel 3.8 Perhitungan pusat cluster 2 iterasi 1	23
Tabel 3.9 Perhitungan fungsi objektif iterasi 1	24
Tabel 3.10 Perubahan matriks partisi baru iterasi 1	24
Tabel 3.11 Perhitungan pusat cluster 1 iterasi 2	25
Tabel 3.12 Perhitungan pusat cluster 2 iterasi 2	26
Tabel 3.13 Perhitungan fungsi objektif iterasi 2	26
Tabel 3.14 Perubahan matriks partisi baru iterasi 2	27
Tabel 3.15 Matriks partisi baru	27
Tabel 3.16 Hasil Pengelompokan	27
Tabel 3.17 Matriks pengelompokan user pada cluster 1	28
Tabel 3.18 Hasil kemiripan setiap <i>user</i> pada cluster 1	28
Tabel 4.1 Spesifikasi Perangkat	30
Tabel 4.2 Skenario Uji Coba Variasi Rank	41
Tabel 4.3 Skenario Uji Coba Variasi <i>Cluster</i>	42
Tabel 4.4 Skenario Uji Coba Variasi Top-m Neighbour	43
Tabel 4.5 Skenario perbandingan menggunakan NDCG	44

DAFTAR KODE PROGRAM

Kode Program 4.1 Input Data Rating Menjadi Matriks R	31
Kode Program 4.2 Proses NMF pada Matriks R.....	32
Kode Program 4.3 Proses Pengelompokan dengan FCM	34
Kode Program 4.4 Menghitung similarity dalam satu cluster dengan PCC	36
Kode Program 4.5 Menentukan top-m neighbour dan memberikan rating	38
Kode Program 4.6 Menentukan Rekomendasi	39
Kode Program 4.7 Mengambil data uji	40
Kode Program 4.8 Evaluasi NDCG	41