

SISTEM REKOMENDASI FILM MENGGUNAKAN METODE *HYBRID* (*CONTENT BASED DAN COLLABORATIVE FILTERING ITEM-BASED*)

Penulis

Triyani Fatchur Rahman 150411100064

Dosen Pembimbing 1

Dr. Noor Ifada, S.T., M.ISD.

19780317 200312 2 001

Dosen Pembimbing 2

Moch. Kautsar Sophan S.Kom., M.MT.

19770713 200212 1 004

ABSTRAK

Sistem rekomendasi adalah sistem yang memprediksi *rating* atau preferensi yang akan diberikan *user* ke suatu *item*. Tujuan dari sistem rekomendasi agar secara otomatis menghasilkan item yang disarankan untuk pengguna sesuai dengan preferensi historis mereka dan menghemat waktu pencarian mereka secara *online* dengan mengumpulkan data yang berhubungan dengan *user* atau *item* yang bermanfaat. Rekomendasi film adalah aplikasi yang paling populer digunakan ditambah dengan *platform* multimedia *online*. Dalam menghasilkan rekomendasi ada beberapa pendekatan metode tradisional yaitu metode *Collaborative Filtering(CF)* dan *Content-Based (CB)*. Meskipun kedua metode memiliki kelebihan, mereka juga memiliki kelemahan tertentu pada masing-masing setiap metode. Secara umum pada pendekatan metode *CF* melaporkan kinerja yang lebih baik daripada pendekatan *CB*, tetapi keberhasilannya bergantung pada jumlah *rating user* yang mencukupi (*Sparsity*). Dalam situasi ini penggunaan pendekatan *CB* muncul sebagai alternatif. Namun demikian, pendekatan ini memiliki keterbatasannya sendiri. Misalnya, kata kunci yang digunakan untuk mewakili konten item mungkin tidak terlalu representatif. Salah satu algoritma yang sering digunakan pada metode *CF* ini yaitu *item-based* dan *user-based*. *Item-based CF* mempunyai akurasi yang lebih baik dibandingkan *user-based CF*. Maka solusi yang ditawarkan dengan menutupi kekurangan kedua metode pada penelitian ini dilakukanlah metode *Hybrid(CB dan CF item-based)*. Pada *CB* akan dilakukan *text mining* untuk melakukan *text preprocessing* nya dan algoritma *K-Means* untuk melakukan *clustering* data karena metode ini telah dikenal sangat efisien dan

diperhitungkan dalam melakukan *clustering*. Dala penelitian ini hasil evaluasi dengan pendekatan metode *hybrid* (*CB* dan *CF item-based*) dalam mengatasi *sparsity* data dan kelemahan pada metode *CB* dan *CF item-based* dalam menghasilkan suatu rekomendasi diperoleh hasil evaluasi DCG dengan menggunakan nilai parameter jumlah $neighbour(Q) = 40$, jumlah $cluster(k) = 520$ dan *Top-10* adalah 0,03840.

Kata Kunci: *Content Based, Collaborative Filtering item-based, K-Means, Sistem Rekomendasi, Text Preprocessing*.

DAFTAR ISI

PENGESAHAN	iii
PERNYATAAN ORISINALITAS	iv
ABSTRAK	v
KATA PENGANTAR	vii
DAFTAR ISI.....	viii
DAFTAR GAMBAR	x
DAFTAR TABEL	xi
DAFTAR PSEUDOCODE	xii
BAB I PENDAHULUAN	13
1.1 Latar Belakang	13
1.2 Perumusan Masalah	14
1.2.1 Permasalahan.....	14
1.2.2 Metode Usulan	14
1.2.3 Pertanyaan Penelitian	14
1.3 Tujuan dan Manfaat.....	15
1.3.1 Tujuan Penelitian	15
1.3.2 Manfaat Penelitian	15
1.4 Batasan Masalah.....	15
1.5 Sistematika Proposal	15
BAB II KAJIAN PUSTAKA	16
2.1 Sistem Rekomendasi	16
2.1.1 <i>Content-based</i>	16
2.1.2 <i>Collaborative filtering</i>	20
2.2 Penelitian Terkait	24
BAB III METODE USULAN.....	28
3.1 Arsitektur Sistem	28
3.2 Metode Evaluasi	37
3.3 Metrik Evaluasi	38
3.4 Tahapan Penelitian	38
3.4.1 Pengumpulan Data	38
3.4.2 Analisis Permasalahan	39
3.4.3 Perancangan Sistem	39

3.4.4	Implementasi	40
3.4.5	Skenario Uji Coba	40
BAB IV	HASIL DAN PEMBAHASAN	42
4.1	Lingkungan Uji Coba	42
4.2	Dataset	43
4.3	Tahapan Pembuatan Program	43
4.3.1	Tahapan berdasarkan IPO	43
4.4	Skenario Pengujian dan Hasil Uji Coba	60
4.4.1	Skenario uji coba <i>cluster</i>	60
4.4.2	Skenario uji coba <i>neighbour</i>	61
4.4.3	Skenario Uji Coba <i>Top N</i>	62
4.5	Tampilan User Interface Sistem	64
BAB V	PENUTUP	66
5.1	Kesimpulan	66
5.2	Saran	66
DAFTAR	PUSTAKA	67
BIOGRAFI	PENULIS	70

DAFTAR GAMBAR

Gambar 2.1 Metode Pada Sistem Rekomendasi	16
Gambar 3.1 Diagram IPO Arsitektur Sistem	28
Gambar 3.2 Metode <i>n-fold cross validation</i>	38
Gambar 4.1 data judulSinopsis.csv	44
Gambar 4.2 Tampilan <i>user interface</i>	64

DAFTAR TABEL

Tabel 2.1 Tabel Penelitian Terkait	27
Tabel 3.1 Hasil Perhitungan <i>TF-IDF</i>	31
Tabel 3.2 Data <i>TF-IDF</i> Judul dan Sinopsis	32
Tabel 3.3 Pemilihan <i>centroid</i>	33
Tabel 3.4 Hasil perhitungan jarak	33
Tabel 3.5 <i>Centroid</i> baru	34
Tabel 3.6 Hasil perhitungan dengan <i>centroid</i> baru	34
Tabel 3.7 <i>Centroid</i> baru kedua.....	34
Tabel 3.8 Hasil <i>Clustering</i>	35
Tabel 3.9 Data rating film	35
Tabel 3.10 Hasil <i>Mean Centered Rating</i>	36
Tabel 3.11 Hasil <i>Similiarity Item</i>	36
Tabel 3.12 Daftar <i>Neighbour</i>	37
Tabel 3.13 Skenario Uji Coba.....	40
Tabel 4.1 Spesifikasi Perangkat	42
Tabel 4.2 Kebutuhan <i>Library</i> Python	42
Tabel 4.3 Data Judul dan Sinopsis	48
Tabel 4.4 Data Kmeans.csv.....	49
Tabel 4.5 Hasil <i>Cluster</i>	52
Tabel 4.6 Data Training	53
Tabel 4.7 Hasil <i>Similiarity</i>	56
Tabel 4.8 Hasil Prediksi Rating	59
Tabel 4.9 Hasil Skenario Uji Coba Jumlah <i>Cluster</i>	60
Tabel 4.10 Hasil Skenario Uji Coba Jumlah <i>Neighbour(Q)</i>	62
Tabel 4.11 Hasil Skenario uji coba <i>Top-N</i>	63

DAFTAR PSEUDOCODE

Pseudocode 4.1 Proses <i>Text Preprocessing</i>	44
Pseudocode 4.2 Proses <i>Case Folding</i>	45
Pseudocode 4.3 Proses mencari jumlah kata dalam sinopsis	45
Pseudocode 4.4 Proses menghitung jumlah kata dalam sinopsis.....	46
Pseudocode 4.5 Proses <i>TF</i>	46
Pseudocode 4.6 Proses Simpan Hasil <i>TF</i> ke <i>dictionary</i>	47
Pseudocode 4.7 Proses <i>IDF</i>	47
Pseudocode 4.8 Proses <i>TF-IDF</i>	48
Pseudocode 4.9 Proses konversi hasil <i>TF-IDF</i> ke format file csv	48
Pseudocode 4.10 Proses Penyesuaian Data <i>TF-IDF</i>	49
Pseudocode 4.11 Proses Memanggil Fungsi Algoritma <i>K-means</i>	50
Pseudocode 4.12 Proses Algoritma <i>K-means</i>	51
Pseudocode 4.13 Proses Input Data <i>Rating</i>	52
Pseudocode 4.14 Proses Konversi Matrik R.....	53
Pseudocode 4.15 Proses <i>Similiarity Item</i>	55
Pseudocode 4.16 Proses Prediksi <i>Rating</i>	58
Pseudocode 4.17 Proses Rekomendasi <i>Top-N</i>	60