

**PENYELESAIAN MASALAH *COLD-START* PADA SISTEM
REKOMENDASI DENGAN MENGGUNAKAN MODEL POPULARITAS
HYBRID BERDASARKAN *CLUSTER USER* DAN *ITEM***

Ummamah
(160411100087)

Pembimbing I : Dr. Noor Ifada S.T., M.ISD.

Pembimbing II : Moch. Kautsar Sophan, S.Kom., M.MT.

ABSTRAK

Cold-start merupakan sebuah masalah yang terjadi pada sistem rekomendasi yang menggunakan pendekatan *memory based*. *Cold-start* terjadi pada *user* baru yang belum memiliki riwayat apapun pada sistem. Kondisi tersebut membuat proses rekomendasi menjadi tidak akurat. Rekomendasi yang diberikan pada *user* baru merupakan *item* yang sering dinilai oleh *user* lainnya. Pada penelitian ini menerapkan metode *Clustering k-modes*, *Fuzzy C-Means* dan *Hybrid* Popularitas Model untuk mengatasi *cold-start*. Algoritma *K-modes* dapat mengelompokkan pengguna berdasarkan atribut umur, jenis kelamin dan pekerjaan. Kemudian, *item* dikelompokkan berdasarkan *rating* menggunakan metode *Fuzzy C-Means Clustering*. Dari hasil uji coba yang telah dilakukan, penerapan *filter rating* pada metode *hybrid* popularity model diperlukan karena dari *hybrid* popularity model memberikan nilai akurasi DCG terbesar pada batas minimal *rating* 2, yaitu 0,8036. Akurasi terbaik dalam teknik *hybrid* popularity model dapat diperoleh dengan menggunakan jumlah *cluster user* $c_x = 10$ untuk metode *K-Modes*, jumlah *cluster item* $c = 95$ untuk metode *Fuzzy C-Means* dan nilai $\alpha = 0$ untuk memastikan bahwa model yang dibangun lebih dipengaruhi oleh *user* populer. Kata kunci : *Cold-start*, *Fuzzy C-means*, *Hybrid* Popularitas Model, *K-modes*, Sistem Rekomendasi

**PENYELESAIAN MASALAH *COLD-START* PADA SISTEM
REKOMENDASI DENGAN MENGGUNAKAN MODEL POPULARITAS
HYBRID BERDASARKAN *CLUSTER USER* DAN *ITEM***

Ummamah
(160411100087)

Pembimbing I : Dr. Noor Ifada S.T., M.ISD.

Pembimbing II : Moch. Kautsar Sophan, S.Kom., M.MT.

ABSTRACT

Cold-start is a problem that occurs in recommendation systems that use a memory based approach. A cold start occurs when a new user doesn't have any history on the system. This condition makes the recommendation process inaccurate. Recommendations given to new users are items that are often rated by other users. This research applies Clustering k-modes, Fuzzy C-Means and Hybrid Popularity Model methods to overcome cold-start. The K-modes algorithm can group users based on attributes of age, sex and occupation. Then, items are grouped based on rating using the Fuzzy C-Means Clustering method. From the results of trials that have been carried out, the application of the filter rating in the hybrid popularity model method is needed because the hybrid popularity model provides the greatest DCG accuracy value at the minimum rating limit of 2, namely 0.8036. The best accuracy in the hybrid popularity model technique can be obtained by using the number of user clusters $cx = 10$ for the K-Modes method, the number of cluster items $c = 95$ for the Fuzzy C-Means method and the value $\alpha = 0$ to ensure that the model built is more influenced by the user. popular.

Keywords: Cold-start, Fuzzy C-means, Hybrid Poularity Model, K-modes, Recommendation System

DAFTAR ISI

HALAMAN PENGESAHAN	ii
HALAMAN PERNYATAAN ORISINALITAS.....	iii
ABSTRAK	iv
KATA PENGANTAR.....	v
DAFTAR ISI	vi
DAFTAR GAMBAR	viii
DAFTAR TABEL.....	ix
DAFTAR ALGORITMA	xi
DAFTAR KODE PROGRAM.....	xii
BAB I PENDAHULUAN	1
1.1. Latar Belakang.....	1
1.2. Perumusan Masalah	2
1.2.1. Permasalahan	3
1.2.2. Usulan Solusi.....	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 Laporan	4
BAB II KAJIAN PUSTAKA.....	6
2.1. Sistem Rekomendasi.....	6
2.2. <i>Item-Based</i>	6
2.3. <i>Cold-Start</i>	7
2.4. <i>K-Modes Clustering</i>	7
2.5. <i>Fuzzy C-Means Clustering</i>	9
2.6. <i>Hybrid</i> Popularitas Model	11
2.7. Penelitian Sebelumnya	12
BAB III METODE USULAN	15
3.1. Data.....	15
3.2. Arsitektur Sistem dan Algoritma	16
3.3. Perhitungan Manual	19
3.3.1. Membangun Matriks Data Demografi User	20

3.3.2.	<i>K-Modes Clustering</i>	20
3.3.3.	Membangun Matriks Data <i>Rating</i>	23
3.3.4.	<i>Fuzzy C-Means</i>	24
3.3.5.	<i>Filter Rating</i>	31
3.3.6.	<i>Hybrid Popularitas Model</i>	32
3.3.7.	Rekomendasi <i>Top-N Item</i>	35
3.4.	Metode Evaluasi	35
3.5.	Skenario Pengujian	36
BAB IV HASIL DAN PEMBAHASAN		38
4.1.	Lingkungan Uji Coba	38
4.2.	Pembuatan Program	38
4.2.1.	Metode	38
4.2.2.	Evaluasi	46
4.3.	Hasil Uji Coba	47
4.3.1.	Skenario 1	47
4.3.2.	Skenario 2	49
4.3.3.	Skenario 3	51
4.3.4.	Skenario 4	52
4.4.	Contoh evaluasi per-user dan hasil pengujian manual	54
BAB V KESIMPULAN DAN SARAN		56
5.1.	Kesimpulan	56
5.2.	Saran	56
REFERENSI		57
LAMPIRAN		61

DAFTAR GAMBAR

Gambar 3.1 Diagram Input Process Output (IPO).....	17
Gambar 3.2 5-Fold Cross Validation	36
Gambar 4.1 GUI hasil rekomendasi hybrid popularity model.....	55

DAFTAR TABEL

<i>Tabel 2.1 Penelitian Sebelumnya.....</i>	<i>13</i>
<i>Tabel 3.1 Contoh data demografi user.....</i>	<i>15</i>
<i>Tabel 3.2 Contoh data rating.....</i>	<i>15</i>
<i>Tabel 3.3 Tabel Konversi Umur [28].....</i>	<i>16</i>
<i>Tabel 3.4 Contoh Hasil Pengelompokan Umur.....</i>	<i>16</i>
<i>Tabel 3.5 Matriks Data Toy Demografi User (Yj).....</i>	<i>20</i>
<i>Tabel 3.6 Inisialisasi K-Modes (cx dan Pt).....</i>	<i>20</i>
<i>Tabel 3.7 Pusat cluster awal K-Modes (Xj).....</i>	<i>20</i>
<i>Tabel 3.8 Pusat cluster awal K-Modes.....</i>	<i>21</i>
<i>Tabel 3.9 Anggota cluster 1 K-Modes.....</i>	<i>21</i>
<i>Tabel 3.10 Anggota cluster 2 K-Modes.....</i>	<i>22</i>
<i>Tabel 3.11 Anggota cluster 3 K-Modes.....</i>	<i>22</i>
<i>Tabel 3.12 Pusat cluster baru K-Modes.....</i>	<i>22</i>
<i>Tabel 3.13 Fungsi objektif K-Modes.....</i>	<i>22</i>
<i>Tabel 3.14 Selisih fungsi objektif K-Modes iterasi ke-1.....</i>	<i>23</i>
<i>Tabel 3.15 Selisih fungsi objektif K-Modes iterasi ke-2.....</i>	<i>23</i>
<i>Tabel 3.16 Hasil Cluster User (CU).....</i>	<i>23</i>
<i>Tabel 3.17 Matriks Data Rating (X).....</i>	<i>24</i>
<i>Tabel 3.18 Inisialisasi Fuzzy C-Means (c, maxIter, w, ε).....</i>	<i>24</i>
<i>Tabel 3.19 Derajat Keanggotaan Acak Fuzzy C-Means (μ_i, k).....</i>	<i>25</i>
<i>Tabel 3.20 Derajat keanggotaan pangkat bobot.....</i>	<i>25</i>
<i>Tabel 3.21 Derajat keanggotaan pangkat bobot dikali rating user 1.....</i>	<i>26</i>
<i>Tabel 3.22 Derajat keanggotaan pangkat bobot dikali rating user 2.....</i>	<i>26</i>
<i>Tabel 3.23 Derajat keanggotaan pangkat bobot dikali rating user 3.....</i>	<i>26</i>
<i>Tabel 3.24 Derajat keanggotaan pangkat bobot dikali rating user 4.....</i>	<i>26</i>
<i>Tabel 3.25 Derajat keanggotaan pangkat bobot dikali rating user 5.....</i>	<i>27</i>
<i>Tabel 3.26 Derajat keanggotaan pangkat bobot dikali rating user 6.....</i>	<i>27</i>
<i>Tabel 3.27 Derajat keanggotaan pangkat bobot dikali rating user 7.....</i>	<i>27</i>
<i>Tabel 3.28 Derajat keanggotaan pangkat bobot dikali rating user 8.....</i>	<i>27</i>
<i>Tabel 3.29 Derajat keanggotaan pangkat bobot dikali rating user 9.....</i>	<i>28</i>
<i>Tabel 3.30 Pusat cluster baru Fuzzy C-Means (vk, j).....</i>	<i>28</i>
<i>Tabel 3.31 Selisih rating dan pusat cluster.....</i>	<i>28</i>
<i>Tabel 3.32 Perubahan matriks partisi Fuzzy C-Means.....</i>	<i>29</i>

<i>Tabel 3.33 Derajat keanggotaan baru Fuzzy C-Means</i>	<i>29</i>
<i>Tabel 3.34 Fungsi objektif Fuzzy C-Means</i>	<i>30</i>
<i>Tabel 3.35 Selisih fungsi objektif Fuzzy C-Means</i>	<i>30</i>
<i>Tabel 3.36 Derajat keanggotaan baru Fuzzy C-Means iterasi ke-26</i>	<i>30</i>
<i>Tabel 3.37 Fungsi objektif Fuzzy C-Means iterasi ke-26.....</i>	<i>31</i>
<i>Tabel 3.38 Selisih fungsi objektif Fuzzy C-Means iterasi ke-26</i>	<i>31</i>
<i>Tabel 3.39 Hasil Cluster Item (CI)</i>	<i>31</i>
<i>Tabel 3.40 Hasil Filter Rating (Y)</i>	<i>32</i>
<i>Tabel 3.41 Inisialisasi hybrid popularitas model</i>	<i>32</i>
<i>Tabel 3.42 Data rating anggota cluster target user (xv, j)</i>	<i>32</i>
<i>Tabel 3.43 Jumlah rating user cluster target terhadap item</i>	<i>33</i>
<i>Tabel 3.44 Normalisasi user popularitas model.....</i>	<i>33</i>
<i>Tabel 3.45 Jumlah rating tiap item cluster terhadap cluster target user</i>	<i>34</i>
<i>Tabel 3.46 Item populer setelah filter IT max</i>	<i>34</i>
<i>Tabel 3.47 Normalisasi item popularitas model.....</i>	<i>34</i>
<i>Tabel 3.48 Hasil hybrid popularitas model.....</i>	<i>35</i>
<i>Tabel 3.49 Hasil rekomendasi Top-N</i>	<i>35</i>
<i>Tabel 3.50 Hasil rekomendasi Top-3</i>	<i>35</i>
<i>Tabel 3.51 Skenario Pengujian.....</i>	<i>37</i>
<i>Tabel 4.1 Perangkat yang digunakan</i>	<i>38</i>
<i>Tabel 4.2 Nilai Parameter Skenario 1.....</i>	<i>48</i>
<i>Tabel 4.3 Hasil evaluasi Skenario 1</i>	<i>49</i>
<i>Tabel 4.4 Nilai Parameter Skenario 2.....</i>	<i>50</i>
<i>Tabel 4.5 Hasil evaluasi Skenario 2</i>	<i>51</i>
<i>Tabel 4.6 Nilai Parameter Skenario 3.....</i>	<i>52</i>
<i>Tabel 4.7 Hasil evaluasi Skenario 3</i>	<i>52</i>
<i>Tabel 4.8 Nilai Parameter Skenario 4.....</i>	<i>53</i>
<i>Tabel 4.9 Hasil evaluasi Skenario 4</i>	<i>53</i>
<i>Tabel 4.10 Nilai rata-rata hybrid 5-fold setiap α untuk 10 item.....</i>	<i>54</i>

DAFTAR ALGORITMA

<i>Algoritma 2.1 K-Modes clustering</i>	<i>9</i>
<i>Algoritma 2.2 Fuzzy C-Means clustering.....</i>	<i>11</i>
<i>Algoritma 3.1 Hybrid Popularitas Model</i>	<i>19</i>

DAFTAR KODE PROGRAM

Kode Program 4.1 Pembuatan data demografi user menjadi matriks.....	39
Kode Program 4.2 K-Modes Clustering	40
Kode Program 4.3 Pembuatan data rating menjadi matriks	41
Kode Program 4.4 Fuzzy C-Means Clustering.....	43
Kode Program 4.5 Filter Rating	43
Kode Program 4.6 Hybrid Popularitas Model.....	45
Kode Program 4.7 Rekomendasi Top-N Item.....	46
Kode Program 4.8 Discounted Comulative Gain (DCG)	47