



Studi literatur penerapan *clustering* data numerik untuk sistem rekomendasi berbasis *collaborative filtering*

Rizki Ashuri Pratama¹, Noor Ifada^{*2}

Email: ¹rizkiashuripratama45@gmail.com, ²noor.ifada@trunojoyo.ac.id

¹²Program Studi Teknik Informatika, Fakultas Teknik, Universitas Trunojoyo Madura

Diterima: 19 Mei 2024 | Direvisi: 09 Agustus 2024 | Disetujui: 21 Agustus 2024

©2020 Program Studi Teknik Informatika Fakultas Ilmu Komputer,
Universitas Muhammadiyah Riau, Indonesia

Abstrak

Sistem rekomendasi membantu pengguna dalam mengatasi masalah untuk menemukan item yang sesuai dengan preferensi dari banyaknya jumlah item yang ada. Sistem rekomendasi memiliki 2 tipe pendekatan: pendekatan berbasis konten dan pendekatan *Collaborative Filtering* (CF). CF dapat dikelompokkan menjadi berbasis model dan memori. Permasalahan yang dihadapi pada metode CF adalah kompleksitas atau waktu komputasi yang lama karena dimensi data yang besar, kelangkaan data dan akurasi. Untuk mengatasi permasalahan tersebut maka digunakan beberapa teknik penambangan data dan pembelajaran mesin untuk kolaborasi dengan CF tradisional. Terdapat banyak penelitian yang menggunakan teknik *clustering* data numerik pada sistem rekomendasi berbasis CF, namun belum ada studi literatur yang dilakukan. Penelitian ini melakukan studi literatur penerapan teknik *clustering* pada data numerik untuk pengembangan metode sistem rekomendasi berbasis pendekatan CF dengan menggunakan 20 buah literatur. Hasilnya, ragam teknik *clustering* yang digunakan dapat dikelompokkan menjadi K-Means, Subspace Clustering, Bi-Clustering, Canopy Clustering, K-Medoids, Evolutionary Heterogeneous Clustering, Fuzzy, Self-Constructing Clustering (SCC) dan Agglomerative Hierarchical Clustering (AHC). Dapat disimpulkan bahwa teknik *clustering* K-Means dan Fuzzy merupakan teknik yang paling banyak ditemukan pada literatur.

Kata kunci: berbasis memori, berbasis model, *clustering*, *collaborative filtering*, studi literatur

Literature review on the implementation of numerical data clustering on collaborative filtering-based recommendation system

Abstract

The recommendation system assists users in finding items that match their preferences from the large number of available items. Recommendation systems have two types of approaches: content-based approach and Collaborative Filtering (CF) approach. CF can be categorized into model-based and memory-based. The problem faced in the CF method is the complexity or long computation time due to the large data dimensions, data sparsity, and accuracy issues. To overcome these problems, several data mining and machine learning techniques are used in collaboration with traditional CF methods. Many studies use numerical data clustering techniques in CF-based recommendation systems, yet no literature review has been done. In this paper, we conduct a literature review on the implementation of clustering techniques on numerical data to develop recommendation system methods based on the CF approach, using 20 related literature. As a result, the various clustering techniques used can be grouped into K-Means, Subspace Clustering, Bi-Clustering, Canopy Clustering, K-Medoids, Evolutionary Heterogeneous Clustering, Fuzzy, Self-Constructing Clustering (SCC), and Agglomerative Hierarchical Clustering (AHC). We find that K-Means and Fuzzy clustering techniques are the most commonly found in the literature.

Keywords: memory-based, model-based, *clustering*, *collaborative filtering*, literature review

1. PENDAHULUAN

Sistem rekomendasi merupakan sebuah program yang menampilkan *item* atau produk tertentu kepada pengguna dengan memprediksi minat pengguna terhadap suatu produk berdasarkan riwayat pengguna dan informasi mengenai produk, dilihat berdasarkan interaksi antara barang dan pengguna [1]. Tujuan utama dari sistem rekomendasi yaitu untuk memprediksi preferensi pengguna melalui sistem yang menawarkan *item* tertentu kepada pengguna di antara sejumlah besar pilihan sesuai dengan selera mereka [2].

Terdapat dua varian pendekatan pada sistem rekomendasi: pendekatan berbasis konten dan pendekatan *Collaborative filtering* (CF). Pendekatan CF dapat dikelompokkan lebih lanjut menjadi CF berbasis model dan berbasis memori. Metode CF berbasis model menggunakan peringkat *item* pengguna untuk mempelajari model prediktif. Sedangkan metode CF memori menggunakan peringkat *item* pengguna yang disimpan dalam sistem untuk memberikan prediksi peringkat *item* baru secara langsung. CF merupakan teknik statistik yang digunakan untuk menemukan pengguna tetangga yang memiliki minat yang sama atau serupa dengan pengguna target, dan memprediksi *item* yang mungkin disukai pengguna target sesuai dengan pengguna tetangga [3]. CF membuat rekomendasi untuk pengguna aktif saat ini dengan menggunakan banyak informasi peringkat historis pengguna tanpa menganalisis konten sumber informasi [4].

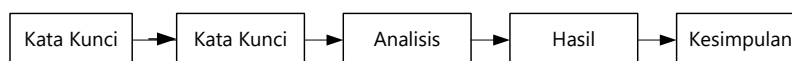
Permasalahan yang dihadapi pada metode CF adalah kompleksitas atau waktu komputasi yang lama karena dimensi data yang besar, kelangkaan data dan akurasi. Dalam mengatasi permasalahan yang terjadi maka digunakan beberapa teknik penambangan data dan pembelajaran mesin sebagai kolaborasi dari metode CF tradisional. Penerapan yang digunakan dapat berupa teknik *clustering* [5, 6]. *Clustering* merupakan metode yang digunakan untuk menemukan struktur pengelompokan dalam kumpulan data yang dicirikan oleh kesamaan terbesar dalam kelompok yang sama dan perbedaan terbesar dalam kelompok yang berbeda [7]. *Clustering* telah banyak digunakan pada berbagai bidang ilmiah seperti: ilmu komputer dan bioinformatika untuk menemukan struktur kelompok dari objek yang direpresentasikan dalam atribut numerik atau kategorik [8].

Terdapat banyak penelitian yang menggunakan teknik *clustering* data numerik pada sistem rekomendasi berbasis CF. Teknik *clustering* data numerik mampu mengatasi permasalahan CF, misalnya akurasi yang lebih baik dibanding CF tradisional [6, 9]. Namun, saat ini masih belum ada studi literatur mengenai penerapan teknik *clustering* pada data numerik untuk pengembangan metode sistem rekomendasi berbasis pendekatan CF.

Oleh karena itu, tujuan penelitian ini adalah melakukan studi literatur mengenai penerapan ragam teknik *clustering* pada data numerik untuk pengembangan metode sistem rekomendasi berbasis pendekatan CF. Pada studi literatur ini akan menggunakan 20 literatur yang berkaitan dengan teknik *clustering* data numerik yang digunakan untuk pengembangan pendekatan CF berbasis konten atau berbasis memori.

2. METODE PENELITIAN

Metode penelitian yang digunakan dalam penelitian ini adalah studi literatur sistematis yang pernah digunakan oleh Masyhur [10]. Seperti yang digambarkan dalam Gambar 1, metode penelitian terdiri dari lima tahapan, yaitu tahap kata kunci, pencarian, analisis, hasil, dan kesimpulan.



Gambar 1. Metode Penelitian

2.1. Kata kunci

Paper yang digunakan pada studi literatur ini dicari dengan menggunakan kata kunci “*clustering algorithm in collaborative filtering recommendations system*”, “*clustering approach in collaborative filtering*”, “*collaborative filtering based on clustering approach*”, “*collaborative filtering dengan teknik clustering*”, atau “*collaborative filtering hybrid clustering*”.

2.2. Pencarian

Pencarian literatur dilakukan dengan memasukkan kata kunci ke dalam situs Google Scholar, IEEE, Springer dan website penyedia jurnal mengenai teknologi informasi lainnya. Literatur yang dipilih berjumlah 20, dimana pemilihan dilakukan berdasarkan kajian singkat yang didapatkan dari judul, abstrak dan metode pada literatur terkait. Beberapa kriteria yang digunakan untuk memilih literatur yang digunakan pada penelitian ini:

1. Literatur terbit dari rentang 2012-2022
2. Literatur merupakan artikel ilmiah yang dipublikasi dan terakreditasi. Kriteria akreditasi untuk publikasi internasional adalah mulai dari ranking Q1 hingga Q3, sedangkan publikasi nasional adalah rentang S1 hingga S3
3. Literatur dari publikasi ilmiah Bahasa Inggris atau Bahasa Indonesia

2.3. Analisis

Tahapan analisis dilakukan pada 20 literatur yang terpilih pada tahap pencarian. Analisis yang dilakukan meliputi identitas literatur, jenis literatur, tujuan, teknik *clustering*, kombinasi metode dan metode pembandingan pada literatur.

2.4. Hasil

Hasil dari analisis setiap literatur akan disajikan dalam bentuk deskripsi, tabel, ataupun gambar dan dikelompokkan berdasarkan dengan jenis analisisnya. Jenis analisisnya yaitu meliputi tujuan, teknik clustering, kombinasi metode, metode pembandingan pada literatur.

2.5. Kesimpulan

Tahap terakhir pada penelitian ini, yaitu dengan membuat kesimpulan berdasarkan hasil analisis studi literatur pada setiap literatur.

3. HASIL DAN PEMBAHASAN

Berdasarkan literatur yang telah didapatkan sesuai dengan tahap satu dan dua, maka ditemukan ragam teknik clustering pada data numerik untuk pengembangan metode sistem rekomendasi berbasis pendekatan CF meliputi K-Means, Fuzzy, Subspace Clustering, Canopy Clustering, K-Medoids, Evolutionary Heterogenous Clustering (EHC), Bi-Clustering, Self-Constructing Clustering (SCC) dan Agglomerative Hierarchical Clustering (AHC). Dari ragam teknik clustering yang ditemukan selanjutnya akan dilakukan perbandingan dan pembahasan lebih lanjut penerapan ragam teknik clustering pada setiap literaturnya.

Hasil analisis akan dikelompokkan berdasarkan jenis analisisnya, yaitu identitas literatur, jenis literatur, tujuan literatur, dataset literatur, target luaran literatur, metode literatur, hasil literatur dan kekurangan literatur.

Tabel 1. Perbandingan Ragam Teknik *Clustering* pada Data Numerik Berbasis Pendekatan CF

Perbandingan	Teknik <i>Clustering</i>							
	K-Means	Fuzzy	Subspace Clustering	Canopy Clustering	K-Medoids	EC	SCC	AHC
Penanganan Dimensi Tinggi	Tidak	Tidak	Ya	Tidak	Tidak	Tidak	Ya	Ya
Penanganan Data Besar	Ya	Ya	Ya	Ya	Tidak	Ya	Ya	Tidak
Penanganan Data Noisy	Tidak	Tidak	Ya	Tidak	Ya	Tidak	Tidak	Ya
Kompleksitas	$O(knt)$	$O(n)$	$O(2n)$	$O(n2)$	$O(k(n-k)2)$	$O(n2)$	$O(nq2)$	$O(n2)$

Tabel 2. Identitas Literatur

No	Literatur	Tahun
1.	Treerattanapitak and Jaruskulchai [11]	2012
2.	Xu [12]	2012
3.	Birtolo and Davide [13]	2013
4.	[14]	2014
5.	Ghazanfar and Prügel-Bennett [15]	2014
6.	Koohi and Kiani [16]	2016
7.	Katarya and Verma [17]	2016
8.	Liao and Lee [18]	2016
9.	Agrawal and Bamnote [9]	2016
10.	Koohi and Kiani [19]	2017
11.	Ifada, et al. [20]	2018
12.	Mirbakhsh and Ling [21]	2018
13.	Chen, et al. [22]	2018
14.	Kant and Mahara [23]	2018
15.	Zarzour [24]	2018
16.	Li, et al. [25]	2019
17.	Deng, et al. [26]	2019
18.	Garanayak, et al. [27]	2019
19.	Ifada, et al. [28]	2020
20.	Andra and Baizal [29]	2022

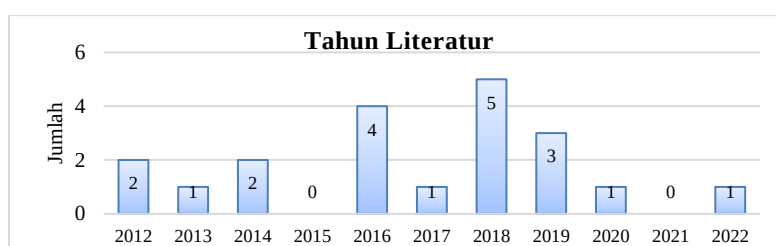
3.1. Perbandingan Ragam Teknik Clustering Pada Data Numerik Berbasis Pendekatan CF

Dilakukan perbandingan ragam teknik *clustering* yang telah ditemukan berdasarkan penanganan dimensi tinggi, penanganan data besar, penanganan data *noisy* dan kompleksitasnya. Hasilnya disajikan pada Tabel 1.

- Penanganan dimensi tinggi: Dimensi (fitur) penting digunakan untuk menganalisis sebuah objek dapat dikelompokkan. Fitur yang banyak dapat berpotensi menimbulkan data *sparse* dan dapat mempengaruhi kualitas pengelompokan
- Penanganan data besar: Ukuran data dapat mempengaruhi kualitas pengelompokan. Beberapa teknik memang hanya dikhususkan untuk menangani data yang kecil dan sebaliknya
- Penanganan data *noisy*: Keberhasilan teknik dalam mengatasi data yang terpaut jauh dengan data lain (*outlier*), karena data *noisy* dapat menyulitkan teknik untuk mengelompokkannya ke kelompok yang sesuai
- Kompleksitas: Perkiraan lama waktu untuk menjalankan atau seberapa efisien teknik *clustering*.

3.2. Identitas Literatur

Identitas literatur yang memuat judul dan tahun terdapat pada Tabel 2. Berdasarkan 20 literatur dilihat dari identitasnya maka dapat diidentifikasi dari tahun dengan rentang tahun 2012 hingga 2022. Pada Gambar 2 literatur paling banyak terdapat pada tahun 2018 yaitu sebanyak 5 literatur dan paling sedikit terdapat pada tahun 2015 dan 2021 karena tidak ditemukan literatur yang memenuhi kriteria.



Gambar 2. Tahun Literatur

3.3. Jenis Literatur

Dari 20 literatur terdapat dua jenis literatur, yaitu jurnal dan proceedings yang ditampilkan pada Tabel 3. Dapat diamati bahwa, dari 20 literatur yang digunakan, didapatkan 75% literatur berbentuk jurnal dan 25% literatur berbentuk *proceedings*.

Jenis literatur sebesar 75% adalah Jurnal yaitu berjumlah 15 literatur. Kualifikasi literatur yang harus dipenuhi adalah Q1 hingga Q3 jika internasional dan S1 hingga S3 jika nasional. Persebaran ranking jurnal ditampilkan pada Tabel 4. Pada Gambar 3 menampilkan visualisasi bahwa dari 15 literatur yang berjenis jurnal prosentase yang paling besar merupakan jurnal dengan ranking Q1 sebesar 73%, selanjutnya Q2 sebesar 13%, Q3 sebesar 7% dan S2 sebesar 7%. Selain itu, tidak terdapat jurnal dengan ranking S1 dan S3 sehingga prosentasenya adalah 0%.

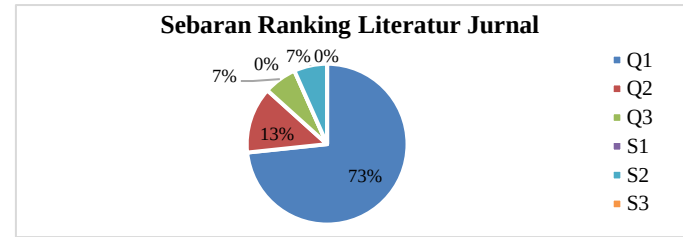
Jenis literatur sebesar 25% adalah Proceedings yaitu sebanyak 5 literatur dengan tiga *publisher* yaitu Association for Computing Machinery (ACM), Springer dan Institute of Electrical and Electronics Engineers (IEEE) dengan masing-masing prosentasenya adalah 20%, 40% dan 40% yang divisualisasikan pada Gambar 4. Persebaran literatur berdasarkan *publisher of proceedings* dimuat pada Tabel 5.

Tabel 3. Jenis Literatur

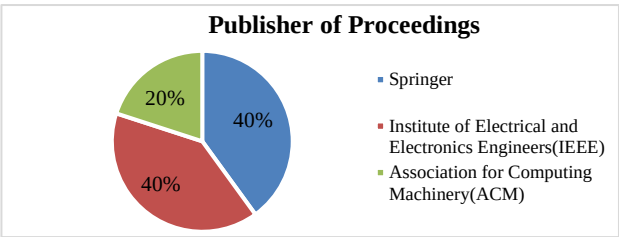
No	Jenis Literatur	Literatur	Jumlah
1.	Jurnal	[11, 13-19, 21-23, 25-27, 29]	15
2.	Proceedings	[9, 12, 20, 24, 28]	5

Tabel 4. Ranking Jurnal

No	Ranking Jurnal	Literatur	Jumlah
1.	Q1	[13-19, 21, 22, 25, 26]	11
2.	Q2	[11, 23]	2
3.	Q3	[27]	1
4.	S1	-	0
5.	S2	[29]	1
6.	S3	-	0



Gambar 3. Sebaran Ranking Literatur Jurnal



Gambar 4. Publisher Proceedings

Tabel 5. Publisher of Proceedings

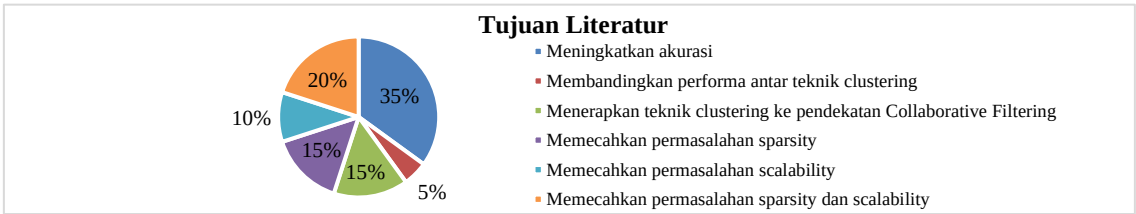
No.	Publisher Proceedings	Literatur	Jumlah
1.	Springer	[9, 24]	2
2.	Institute of Electrical and Electronics Engineers (IEEE)	[20, 28]	2
3.	Association for Computing Machinery (ACM)	[12]	1

3.4. Tujuan Literatur

Ditinjau dari tujuannya, 20 literatur dibagi menjadi 6 bagian dan persebaran literatur dimuat pada Tabel 6. Berdasarkan visualiasi Gambar 5, didapatkan fokus tujuan literatur dalam menggunakan teknik clustering pada pendekatan CF yang paling banyak adalah 35%, yaitu berfokus pada meningkatkan akurasi sistem rekomendasi, namun apabila tujuan memecahkan permasalahan tidak dispesifikkan maka tujuan yang paling besar adalah mengatasi permasalahan yang terjadi pada CF yang mencapai 45%.

Tabel 6. Tujuan Literatur

No	Tujuan	Literatur	Jumlah
1.	Meningkatkan akurasi	[11, 13, 14, 17, 21, 24, 26]	7
2.	Membandingkan performa antar teknik clustering	[28]	1
3.	Menerapkan teknik clustering ke pendekatan Collaborative Filtering	[16, 22, 27]	3
4.	Memecahkan permasalahan sparsity	[18, 23, 29]	3
5.	Memecahkan permasalahan scalability	[9, 15]	2
6.	Memecahkan permasalahan sparsity dan scalability	[12, 19, 20, 25]	4



Gambar 5. Tujuan Literatur

3.5. Dataset Literatur

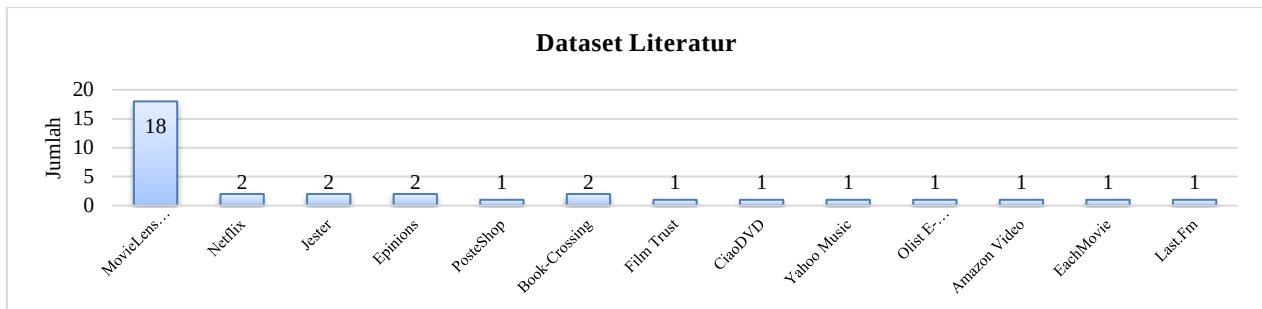
Tabel 7 menunjukan dataset yang digunakan pada literatur, terdapat literatur yang menggunakan dataset lebih dari satu. Dataset yang paling banyak digunakan pada Collaborative Filtering dengan teknik clustering pada data numerik adalah MovieLens dari GroupLens yaitu sebanyak 18 literatur. Dataset yang minim digunakan yaitu dataset Netflix, Jester, Epinions dan Book-Crossing masing-masing digunakan pada 2 literatur. Sedangkan pada dataset PosteShop, Film Trust, CiaoDVD, Yahoo Music, Olist E-Commerce, Amazon Video, EachMovie dan Last.Fm masing-masing hanya digunakan pada 1 literatur. Visualisasi dari penggunaan dataset terdapat pada Gambar 6.

K-Fold Cross Validation dapat digunakan untuk meningkatkan akurasi pada sistem rekomendasi, sehingga dilakukan analisis terhadap penggunaannya. Gambar 7 dan Gambar 8 menunjukkan bahwa $K = 5$ paling umum digunakan pada sistem rekomendasi dengan pendekatan CF dengan splitting data paling banyak digunakan adalah 80% data training dan 20% data test.

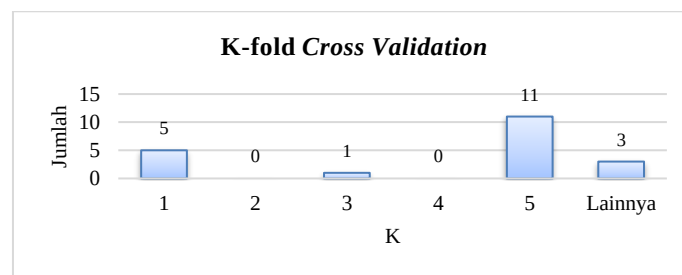
Tabel 7. Dataset Literatur

No	Dataset	Literatur	Jumlah
1.	MovieLens GroupLens	[9, 11-19, 21-23, 25-28]	18
2.	Netflix	[21, 24]	2
3.	Jester	[13, 19]	2

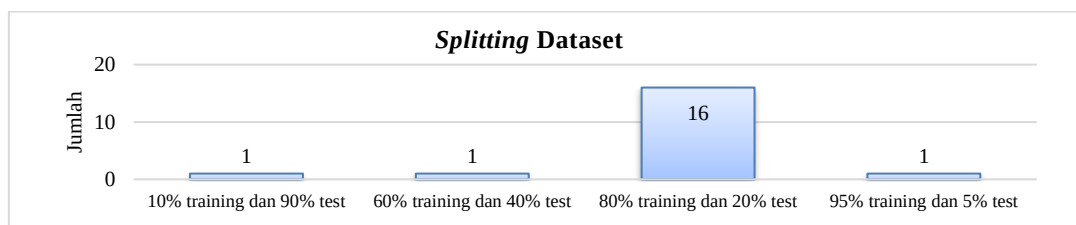
4.	Epinions	[13, 18]	2
5.	PosteShop	[13]	1
6.	Book-Croosing	[14, 18]	2
7.	Film Trust	[15]	1
8.	CiaoDVD	[22]	1
9.	Yahoo Music	[26]	1
10.	Olist E-Commerce	[29]	1
11.	Amazon Video	[18]	1
12.	EachMovie	[23]	1
13.	Last.FM	[12]	1



Gambar 6. Dataset Literatur



Gambar 7. Penggunaan K-Fold Cross Validation



Gambar 8. Splitting Dataset

3.6. Target Luaran Literatur

Tabel 8 menampilkan target luaran literatur yang dikelompokkan menjadi tiga, yaitu prediksi rating, rekomendasi Top-N serta prediksi rating dan rekomendasi Top-N. Berdasarkan Gambar 9 maka target luaran urut dari yang paling besar adalah prediksi rating yaitu sebanyak 12 literatur, rekomendasi Top-N sebanyak 7 literatur serta prediksi rating dan rekomendasi Top-N sebanyak 1 literatur.

Berdasarkan tiga kelompok target luaran, target luaran menggunakan metrik evaluasinya masing-masing. Target luaran prediksi rating menggunakan metrik evaluasi prediksi yang dimuat pada Tabel 9. Target luaran rekomendasi Top-N menggunakan metrik evaluasi rekomendasi yang dimuat pada Tabel 10. Sedangkan, target luaran prediksi rating dan rekomendasi Top-N menggunakan metrik evaluasi prediksi dan metrik evaluasi rekomendasi. Dari target luaran prediksi rating sebanyak 12 literatur serta target luaran prediksi dan rekomendasi sebanyak 1 literatur. Berdasarkan visualisasi Gambar 10, maka metrik evaluasi prediksi yang paling banyak digunakan adalah metrik MAE dan RMSE yang digunakan secara bersamaan. Penggunaan metrik evaluasi rekomendasi ditampilkan pada Gambar 11 memiliki enam pengkondisian, metrik yang paling banyak digunakan adalah *accuracy*, *precision* dan *recall* yang digunakan secara bersamaan serta *precision*, *recall* dan *f1-score* yang digunakan secara bersamaan.

Tabel 8. Target Luaran Literatur

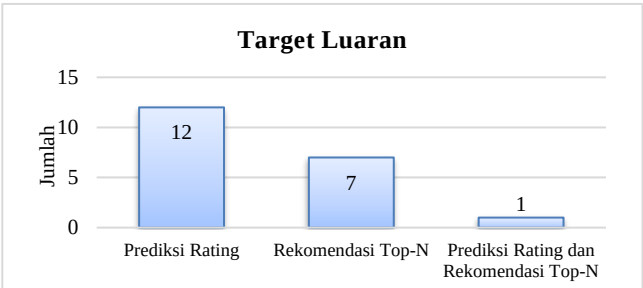
No	Target Luaran	Literatur	Jumlah
1.	Prediksi Rating	[11, 13, 15, 17, 18, 21-25, 27, 29]	12
2.	Rekomendasi Top-N	[9, 12, 14, 16, 19, 20, 28]	7
3.	Prediksi Rating dan Rekomendasi Top-N	[27]	1

Tabel 9. Literatur berdasarkan Metrik Evaluasi Prediksi Rating

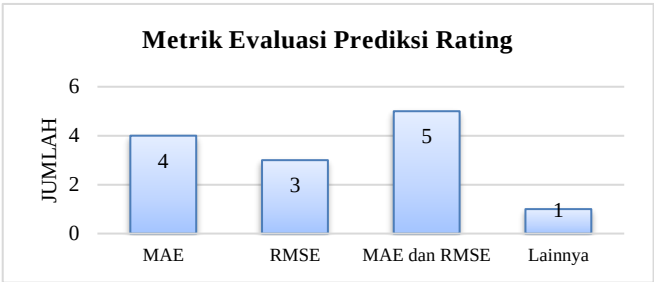
No	Metrik Evaluasi Prediksi	Literatur	Jumlah
1.	MAE	[11, 15, 17, 22]	4
2.	RMSE	[13, 21, 25]	3
3.	MAE dan RMSE	[18, 23, 24, 26, 29]	5
4.	Lainnya	[27]	1

Tabel 10. Literatur berdasarkan Metrik Evaluasi Rekomendasi Top-N

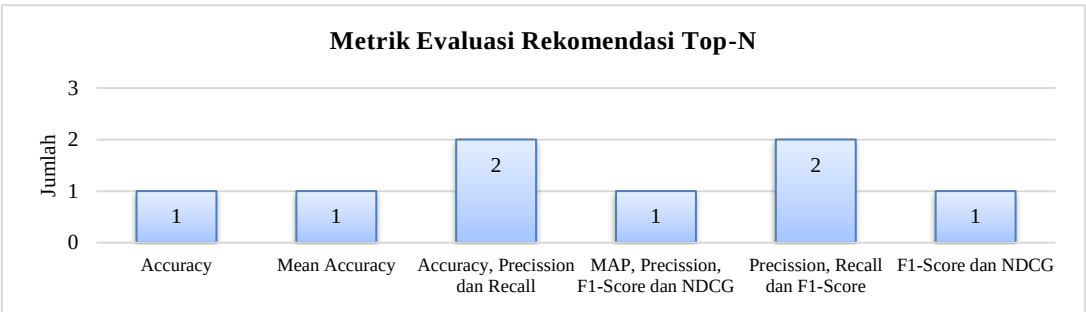
No	Metrik Evaluasi Rekomendasi Top-N	Literatur	Jumlah
1.	Accuracy	[9]	1
2.	Mean Accuracy	[14]	1
3.	Accuracy, Precision dan Recall	[16, 19]	2
4.	MAP, Precision, F1-Score dan NDCG	[12]	1
5.	Precision, Recall dan F1-Score	[20, 26]	2
6.	F1-Score dan NDCG	[28]	1



Gambar 9. Target Luaran



Gambar 10. Metrik Evaluasi Prediksi Rating



Gambar 11. Metrik Evaluasi Rekomendasi Top-N

3.7. Kombinasi dan Metode Pembanding Teknik Clustering

Setiap literaturnya menggunakan teknik *clustering* dalam kombinasi yang berbeda-beda dan dibandingkan dengan metode yang berbeda-beda pula. Tabel 11 memuat kombinasi dan metode pembanding K-Means, Tabel 12 memuat kombinasi dan metode pembanding Fuzzy. Sedangkan Tabel 13 memuat kombinasi dan metode pembanding teknik *clustering* lain.

Tabel 11. Kombinasi dan Metode Pembanding K-Means

No	Kombinasi K-Means	Metode Pembanding K-Means	
1.	Hybrid (Content Based+Collaboration Filtering) dengan susunan Clustering (KMeansProbPower) guna memisahkan grey -sheep ke kelompok tersendiri dimana CF untuk menangani non grey-sheep dan	Clustering	KMeansPlus
			KMeansPlusPower
			KMeansPlusProbPower

	Content-Based (Klasifikasi/Regresi pada informasi teks film(SVMReg)) untuk menangani grey-sheep	Ukuran Jarak	Pearson Correlation (PCC)
			Pearson Correlation with Default Votes (PCCDV)
			Vector Similarity (VS)
			Vector Similarity with Default Votes (VSDV)
		Algoritma Text Categorization	K Nearest Neighbours (KNN)
			Naive Bayes classifier (NB)
			Decision tree (C4.5)
			Support Vector Machines Classification (SVMClass)
2.	Clustering+PSO+Prediksi Rating (K-Means+PSO+Fuzzy C-Means)	Self-Organizing Map(SOM)	
		K-Means	
		Genetic Algorithm+K-Means(GA-KM)	
		Principal Component Analysis (PCA) GAKM	
		PCA SOM	
		PCA KM	
3.	Clustering-Based Matrix Factorization (Matriks Faktorisasi (MF)+Clustering (K-Means)	Expectation Maximization (EM)	
	CB-AsydsVD (Matriks Faktorisasi(MF)+Clustering (K-Means)+SVD Asimetris)	Density-Based K-Means	
		Farthest First	
4.	Clustering (K-Means)+CF User-Based	User-Based Collaborative Filtering	
	Clustering (K-Means)+Matriks Faktorisasi (PCA)+CF User-Based		
5.	User-Item Preference Category Matrix+Canopy Clustering+K-Means Clustering+Similaritas user+CF User-Based	Collaborative Filtering tradisional	
6.	Clustering (Item K-Means)+Similaritas (Pearson Correlation)+CF Item Based	-	
7.	Matriks Faktorisasi (NMF, SVD dan PCA)+Clustering (K-Means dan Fuzzy C-Means)+CF User-Based	Non-Negative Matrix Factorization (NMF)	
		Principal Component Analysis (PCA)	
		Singular Value Deomposition (SVD)	
8.	Matriks Faktorisasi (PCA)+Clustering (K-Means)+Penanganan Sparsity (Imputasi rata-rata nilai pengguna)+CF User-Based	K-Means CF	
		Hierarchical Clustering	

Tabel 12. Kombinasi dan Pembanding Fuzzy

No	Kombinasi Fuzzy	Metode Pembanding Fuzzy
1.	Clustering+Collaborative Filtering (Exponential Fuzzy C-Means (XFCM)+Item-Based)	Item-Based Collaborative Filtering
		Fuzzy C-Means
		Fuzzy C-Means with Entropy (FCME)
		Singular Value Decomposition (SVD)
2.	Reduksi Dimensi+Clustering (Fuzzy C-Means)+Prediksi SubGroup(digabung)	User-Based
		Item-Based
		Personality Diagnosis (PD)
		Singular Value Decomposition (SVD)
		Non-negative Matrix Factorization (NMF)
		Maximum Margin Matrix Factorization
		NPCA
		POP
3.	Clustering+Collaborative Filtering (Fuzzy C-Means+User/ItemBased (IFCCF)	K-Means
	Clustering+User/Item Information+Collaborative Filtering (Fuzzy C-Means+User/ItemBased (IFCCF) dan Fuzzy C-Means+Trust+User/ItemBased (TRACCF)	Self-Organizing Map (SOM)
		Slope One(CF Tradisional)
4.	Fuzzy+Collaborative Filtering (User-Based)	Fuzzy Collaborative Fitering milik Lucas
		Fuzzy Collaborative Fitering milik Zenebe dan Norcio

5.	Clustering (Fuzzy C-Means)+Defuzzifikasi (Maximum dan Center Of Gravity (COG))+Collaboration Filtering User-Based (Pearson dan Average)	K-Means
		Self-Organizing Map(SOM)
6.	Sparsity Removal (Relative Weighting (Adjusted Cosine+Jaccard))+Clustering (Fuzzy C-Means)+CF Item-Based	Item-Based Collaborative Filtering
7.	Matriks Faktorisasi(NMF, SVD dan PCA)+Clustering (K-Means dan Fuzzy C-Means)+CF User-Based	Non-Negative Matrix Factorization (NMF)
		Principal Component Analysis (PCA)
		Singular Value Deomposition (SVD)
8.	Clustering+PSO+Prediksi Rating (K-Means+PSO+Fuzzy C-Means)	Self-Organizing Map (SOM)
		K-Means
		Genetic Algorithm+K-Means (GA-KM)
		Principal Component Analysis (PCA) GAKM
		PCA SOM
		PCA KM

Tabel 13. Kombinasi dan Pembanding Subspace Clustering, Bi-Clustering, Canopy Clustering, K-Medoids, EHC, SCC dan AHC

No	Teknik Clustering	Kombinasi	Metode Pembanding
1.	Subspace Clustering	Data Representation+Constructing Item Subspace+Remove Redudancy+(Neighbour User+Finding by Using+Neighbor Trees)+Estimating Rating Value (Average Pred dan Sim Pred)+Recommendation	K-Means
			Pearson Correlation (PCC)
			Non-Negative Matrix Factorization model for clustering (NNMF)
2.	Bi-Clustering	Clustering (BiClustering)+CF User-Based+CF Item-Based+Kombinasi CF User-Based dan Item-Based	RSVD
			Item-Rank
			CF Item-Based
			CF User-Based
			LDA
			iExpand
3.	Canopy Clustering	User-Item Preference Category Matrix+Canopy Clustering+K-Means Clustering+Similaritas user+CF User-Based	Artificial Immune Systems Based Collaborative Filtering (AIS+CF)
			Collaborative Filtering tradisional
			Pearson Correlation based CF (PCCF)
			New Heuristic Similarity Model (NHSM)
4.	K-Medoids	Similaritas item (Divergensi KullBack-Leibler)+Clustering (K-Medoids)+CF Item-Based	Combining Jaccard and MSD (JMSD)
			Bhattacharyya Coefficient based CF (BCF)
5.	EHC	Evolutionary Heterogeneous Clustering (EHC)+Collaborative Filtering User-Based	Collaborative Filtering tradisional
			K-Means+Collaborative Filtering
6.	SCC	User Labeling (SCC)+Dimensionality Reduction (SCC)+Korelasi Graph+Random Walks (Rekomendasi Group)+Re-Transformasi (Rekomendasi Individu)	ItemRank
			iExpand(LDA)
			PRM2(SVD)
			BiFU(K-Means)
			ICRRS
7.	AHC	Clustering (AHC)+CF	Hybrid
			Content-Based

4. KESIMPULAN

Setelah dilakukan studi literatur pada 20 literatur yang telah diseleksi sesuai dengan kriteria maka didapatkan informasi seputar penggunaan teknik clustering data numerik pada pendekatan CF. Pertama, literatur yang paling banyak mengimplementasikan teknik clustering ke pendekatan CF merupakan tahun 2018 yaitu sebanyak 5 literatur. Kedua, jenis literatur yang didapatkan didominasi dengan jurnal internasional dengan ranking Q1. Ketiga, fokus tujuan literatur terbesar adalah 45%, yaitu menggunakan teknik *clustering* ke pendekatan CF untuk memecahkan permasalahan yang terjadi pada pendekatan CF seperti masalah *sparsity*, *scalability*, dan *sparsity* sekaligus *scalability*. Keempat, *dataset* yang banyak digunakan untuk evaluasi teknik *clustering* pada pendekatan CF merupakan MovieLens GroupLens. Kelima, target luaran literatur didominasi untuk menghasilkan prediksi rating dibandingkan dengan rekomendasi Top-N ataupun prediksi rating dan rekomendasi Top-N

sekalius. Keenam, ragam teknik *clustering* yang digunakan untuk pengembangan metode sistem rekomendasi berbasis pendekatan CF dapat dikelompokkan menjadi K-Means, Subspace Clustering, Bi-Clustering, Canopy Clustering, K-Medoids, Evolutionary Heterogeneous Clustering (EHC), Fuzzy, Self-Constructing Clustering (SCC) dan Agglomerative Hierarchical Clustering (AHC). Ketujuh, K-Means dan Fuzzy merupakan teknik *clustering* yang paling banyak digunakan dengan jumlah literatur yang menggunakannya adalah 8 literatur.

Meskipun teknik *clustering* K-Means dan Fuzzy paling banyak digunakan pada 20 literatur, namun ada hal yang perlu diperhatikan yaitu pada kedua teknik tersebut memerlukan inisialisasi jumlah *cluster K* dimana setiap *K* memiliki akurasi yang berbeda dan perlu dilakukan analisis *K* terbaik untuk model yang dibangun. Terdapat literatur yang mengatasi permasalahan pemilihan *K* terbaik, yaitu dengan melakukan kombinasi teknik *clustering* K-Means, Particle Swarm Optimization (SOM) dan Fuzzy C-Means (FCM).

5. DAFTAR PUSTAKA

- [1] Patel, K. and Patel, H. B., 2020, A state-of-the-art survey on recommendation system and prospective extensions. *Computers and Electronics in Agriculture*, 178, p. 105779, doi: <https://doi.org/10.1016/j.compag.2020.105779>.
- [2] Afoudi, Y., Lazaar, M., and Al Achhab, M., 2021, Hybrid recommendation system combined content-based filtering and collaborative prediction using artificial neural network. *Simulation Modelling Practice and Theory*, 113, p. 102375, doi: <https://doi.org/10.1016/j.simpat.2021.102375>.
- [3] Juan, W., Yue-xin, L., and Chun-ying, W., 2019, Survey of recommendation based on collaborative filtering. *Journal of Physics: Conference Series*, 1314(1), p. 012078, doi: <https://doi.org/10.1088/1742-6596/1314/1/012078>.
- [4] Chen, R., Hua, Q., Chang, Y.-S., Wang, B., Zhang, L., and Kong, X., 2018, A survey of collaborative filtering-based recommender systems: From traditional methods to hybrid methods based on social networks. *IEEE Access*, 6, pp. 64301-64320, doi: <https://doi.org/10.1109/ACCESS.2018.2877208>.
- [5] Chen, J., Zhao, C., Uliji, and Chen, L., 2020, Collaborative filtering recommendation algorithm based on user correlation and evolutionary clustering. *Complex & Intelligent Systems* 6, pp. 147-156, doi: <https://doi.org/10.1007/s40747-019-00123-5>.
- [6] Son, N. T., Dat, D. H., Trung, N. Q., and Anh, B. N., 2017, Combination of dimensionality reduction and user clustering for collaborative-filtering. *The 2017 International Conference on Computer Science and Artificial Intelligence (CSAI /17)*, Jakarta, Indonesia, pp. 125-130, doi: <https://doi.org/10.1145/3168390.3168405>.
- [7] Sinaga, K. P. and Yang, M.-S., 2020, Unsupervised K-means clustering algorithm. *IEEE Access*, 8, pp. 80716-80727, doi: <https://doi.org/10.1109/ACCESS.2020.2988796>.
- [8] Cheung, Y.-m. and Jia, H., 2013, Categorical-and-numerical-attribute data clustering based on a unified similarity metric without knowing cluster number. *Pattern Recognition*, 46(8), pp. 2228-2238, doi: <https://doi.org/10.1016/j.patcog.2013.01.027>.
- [9] Agrawal, S. S. and Bamnote, G. R., 2016, Implementing and evaluating collaborative filtering (CF) using clustering. *The First International Conference on Information and Communication Technology for Intelligent Systems*, Ahmedabad, India, pp. 153-163, doi: https://doi.org/10.1007/978-3-319-30927-9_16.
- [10] Masyhur, F., 2017, Penelitian e-Government di Indonesia: Studi Literatur Sistematis dari Perspektif Dimensi Pemeringkatan e-Government Indonesia (PeGI). *Jurnal IPTEKKOM (Jurnal Ilmu Pengetahuan & Teknologi Informasi)*, 19(1), pp. 51-62, doi: <https://doi.org/10.17933/iptekkom.19.1.2017.51-62>.
- [11] Treerattanapitak, K. and Jaruskulchai, C., 2012, Exponential fuzzy C-means for collaborative filtering. *Journal of Computer Science and Technology*, 27, pp. 567-576, doi: <https://doi.org/10.1007/s11390-012-1244-x>.
- [12] Xu, B., Jiajun Bu, Chun Chen, and Deng Cai, 2012, An exploration of improving collaborative recommender systems via user-item subgroups. *The 21st International Conference on World Wide Web (WWW /12)*, Lyon, France, pp. 21-30, doi: <https://dl.acm.org/doi/abs/10.1145/2187836.2187840>.
- [13] Birtolo, C. and Davide, R., 2013, Advances in clustering collaborative filtering by means of fuzzy C-means and trust. *Expert Systems with Applications*, 40(17), pp. 6997-7009, doi: <https://doi.org/10.1016/j.eswa.2013.06.022>.
- [14] Son, L. H., 2014, HU-FCF: a hybrid user-based fuzzy collaborative filtering method in recommender systems. *Expert Systems with Applications: An International Journal*, 41(15), pp. 6861-6870, doi: <https://doi.org/10.1016/j.eswa.2014.05.001>.
- [15] Ghazanfar, M. A. and Prügel-Bennett, A., 2014, Leveraging clustering approaches to solve the gray-sheep users problem in recommender systems. *Expert Systems with Applications*, 41(7), pp. 3261-3275, doi: <https://doi.org/10.1016/j.eswa.2013.11.010>.
- [16] Koohi, H. and Kiani, K., 2016, User based Collaborative Filtering using fuzzy C-means. *Measurement*, 91, pp. 134-139, doi: <https://doi.org/10.1016/j.measurement.2016.05.058>.
- [17] Katarya, R. and Verma, O. P., 2016, A collaborative recommender system enhanced with particle swarm optimization technique. *Multimedia Tools and Applications*, 75, pp. 9225-9239, doi: <https://doi.org/10.1007/s11042-016-3481-4>.
- [18] Liao, C.-L. and Lee, S.-J., 2016, A clustering based approach to improving the efficiency of collaborative filtering recommendation. *Electronic Commerce Research and Applications*, 18, pp. 1-9, doi: <https://doi.org/10.1016/j.elerap.2016.05.001>.
- [19] Koohi, H. and Kiani, K., 2017, A new method to find neighbor users that improves the performance of collaborative filtering. *Expert Systems with Applications*, 83, pp. 30-39, doi: <https://doi.org/10.1016/j.eswa.2017.04.027>.
- [20] Ifada, N., Prasetyo, E. H., and Mula'ab, 2018, Employing sparsity removal approach and Fuzzy C-Means clustering technique on a movie recommendation system. *The 2018 International Conference on Computer Engineering, Network and Intelligent Multimedia (CENIM)*, Surabaya, Indonesia, pp. 157-162, doi: <https://doi.org/10.1109/CENIM.2018.8711270>.
- [21] Mirbakhsh, N. and Ling, C. X., 2018, Leveraging clustering to improve collaborative filtering. *Information Systems Frontiers*, 20, pp. 111-124, doi: <https://doi.org/10.1007/s10796-016-9668-4>.
- [22] Chen, J., Uliji, Wang, H., and Yan, Z., 2018, Evolutionary heterogeneous clustering for rating prediction based on user collaborative filtering. *Swarm and Evolutionary Computation*, 38, pp. 35-41, doi: <https://doi.org/10.1016/j.swevo.2017.05.008>.
- [23] Kant, S. and Mahara, T., 2018, Merging user and item based collaborative filtering to alleviate data sparsity. *International Journal of System Assurance Engineering and Management*, 9, pp. 173-179, doi: <https://doi.org/10.1007/s13198-016-0500-9>.
- [24] Zarzour, H., Faiz Maazouzi, Mohamed Soltani, and Chaouki Chemam, 2018, An improved collaborative filtering recommendation algorithm for big data. *The 6th IFIP International Conference on Computational Intelligence and Its Applications (CIIA 2018)*, Oran, Algeria, pp. 660-668, doi: https://doi.org/10.1007/978-3-319-89743-1_56.
- [25] Li, J., Zhang, K., Yang, X., Wei, P., Wang, J., Mitra, K., and Ranjan, R., 2019, Category preferred canopy-K-means based collaborative filtering algorithm. *Future Generation Computer Systems*, 93, pp. 1046-1054, doi: <https://doi.org/10.1016/j.future.2018.04.025>.
- [26] Deng, J., Guo, J., and Wang, Y., 2019, A Novel K-medoids clustering recommendation algorithm based on probability distribution for collaborative filtering. *Knowledge-Based Systems*, 175, pp. 96-106, doi: <https://doi.org/10.1016/j.knsys.2019.03.009>.
- [27] Garanyak, M., Mohanty, S. N., Jagadev, A. K., and Sahoo, S., 2019, Recommender system using item based collaborative filtering (CF) and K-means. *International Journal of Knowledge-based and Intelligent Engineering Systems*, 23(2), pp. 93-101, doi: <https://doi.org/10.3233/KES-190402>.

- [28] Ifada, N., Fitriantama, M. N., Sophan, M. K., and Wahyuni, S., 2020, Collaborative Filtering Item Recommendation Methods based on Matrix Factorization and Clustering Approaches. *The 10th Electrical Power, Electronics, Communications, Controls and Informatics Seminar (EECCIS)*, Malang (Online), pp. 226-230, doi: <https://doi.org/10.1109/EECCIS49483.2020.9263450>.
- [29] Andra, D. and Baizal, A., 2022, E-commerce Recommender System Using PCA and K-Means Clustering. *Jurnal RESTI (Rekayasa Sistem dan Teknologi Informasi)*, 6(1), pp. 57-63, doi: <https://doi.org/10.29207/resti.v6i1.3782>.