

**PERBANDINGAN ALGORITMA MODEL-BASED COLLABORATIVE  
FILTERING DALAM MENGATASI DATA SPARSITY**

**SKRIPSI**



disusun oleh

**Fariz Aditama**

**17.11.0948**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2021**

**PERBANDINGAN ALGORITMA MODEL-BASED COLLABORATIVE  
FILTERING DALAM MENGATASI DATA SPARSITY**

**SKRIPSI**

Untuk memenuhi sebagian persyaratan  
mencapai gelar Sarjana  
pada Program Studi Informatika



disusun oleh

**Fariz Aditama**

**17.11.0948**

**PROGRAM SARJANA  
PROGRAM STUDI INFORMATIKA  
FAKULTAS ILMU KOMPUTER  
UNIVERSITAS AMIKOM YOGYAKARTA  
YOGYAKARTA  
2021**

**PERSETUJUAN**

**SKRIPSI**

**PERBANDINGAN ALGORITMA MODEL-BASED COLLABORATIVE  
FILTERING DALAM MENGATASI DATA SPARSITY**

yang dipersiapkan dan disusun oleh

**Fariz Aditama**

**17.11.0948**

telah disetujui oleh Dosen Pembimbing Skripsi  
pada tanggal 4 Februari 2021

**Dosen Pembimbing,**

**Hartatik, S.T., M.Cs.**

**NIK. 190302232**

**PENGESAHAN****SKRIPSI****PERBANDINGAN ALGORITMA MODEL-BASED COLLABORATIVE  
FILTERING DALAM MENGATASI DATA SPARSITY**

yang dipersiapkan dan disusun oleh

**Fariz Aditama**

**17.11.0948**

telah dipertahankan di depan Dewan Penguji  
pada tanggal 17 Februari 2021

**Susunan Dewan Penguji**

**Nama Penguji**

**Tanda Tangan**

**Erni Seniwati, S.Kom, M.Cs**  
**NIK. 190302187**

**Rumini, M.Kom**  
**NIK. 190302246**

**Hartatik, S.T., M.Cs**  
**NIK. 190302232**

Skripsi ini telah diterima sebagai salah satu persyaratan  
untuk memperoleh gelar Sarjana Komputer  
Tanggal 17 Februari 2021

**DEKAN FAKULTAS ILMU KOMPUTER**

**Krisnawati, S.Si, M.T.**  
**NIK. 190302038**

## PERNYATAAN

Saya yang bertandatangan di bawah ini menyatakan bahwa skripsi ini merupakan karya saya sendiri (ASLI) dan isi pada skripsi ini tidak terdapat karya yang pernah diajukan oleh orang lain untuk memperoleh gelar akademis di suatu institusi pendidikan tinggi manapun, dan sepanjang pengetahuan saya juga tidak terdapat karya atau pendapat yang pernah ditulis dan/atau diterbitkan oleh orang lain, kecuali yang secara tertulis diacu dalam naskah ini dan disebutkan dalam daftar pustaka.

Segala sesuatu yang terkait dengan masalah dan karya yang telah dibuat adalah menjadi tanggungjawab saya pribadi.

Yogyakarta, 3 Maret 2021



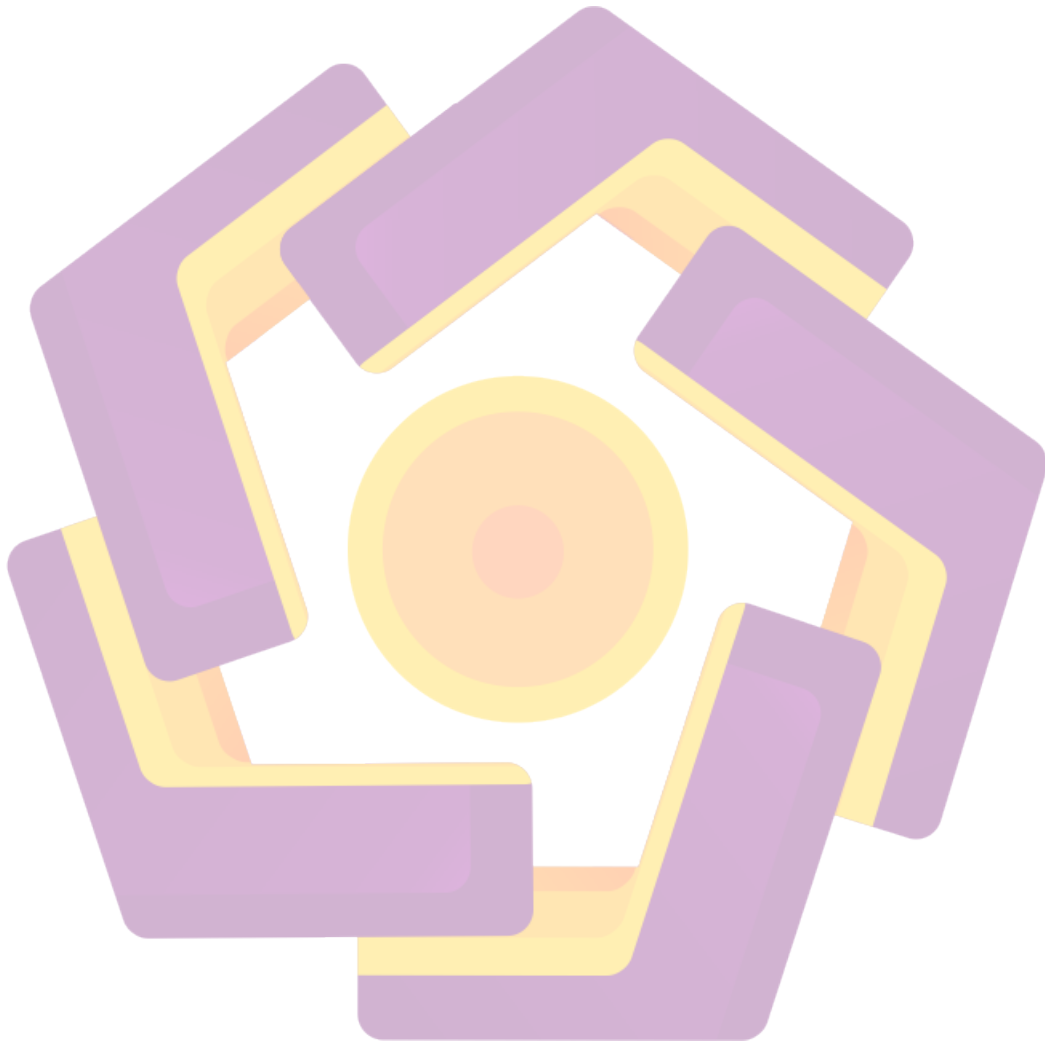
Fariz Aditama

NIM 17.11.0948

**MOTTO**

*“Life isn’t about finding yourself. Life is about creating yourself.”*

-George Bernard Shaw



## PERSEMBAHAN

Puji syukur saya ucapkan kehadiran Allah SWT yang telah melimpahkan rahmat dan karunia-Nya sehingga saya dapat menyelesaikan penyusunan skripsi ini. Dengan ini saya mempersembahkan skripsi ini kepada semua pihak yang terlibat secara langsung atau tidak langsung yaitu untuk :

1. Kedua orang tua saya (Achmad Mudjito dan Dwi Roselawaty).
2. Kedua kakak saya (Kartika Nur Gemilang dan Rizkiawan Nur Latifi).
3. Dosen pembimbing saya Ibu Hartatik. S.T., M.Cs.
4. Dosen Universitas AMIKOM Yogyakarta yang telah memberikan banyak segala ilmu selama saya berkuliah.
5. Teman sepenelitian (Bangdim, Anjas, Devita, Remu) yang selalu membantu memberikan dukungan kepada saya untuk dapat menyelesaikan skripsi.
6. Teman yang lebih yang selalu mengingatkan dan memberikan semangat untuk menyelesaikan skripsi (Pawiloy)
7. Teman sekelas 17-S11F-01 yang telah memberikan dukungan selama perkuliahan.
8. Teman bermain yang setia menemani saat pusing (Adin, Bumi, Kecer, Tazam)

## KATA PENGANTAR

Penulis panjatkan puji dan syukur kehadirat Tuhan Yang Maha Esa atas berkat dan rahmat-Nya sehingga penulis dapat menyelesaikan skripsi dengan baik yang berjudul **“PERBANDINGAN ALGORITMA MODEL-BASED COLLABORATIVE FILTERING DALAM MENGATASI DATA SPARSITY”** disusun sebagai salah satu syarat utama untuk menyelesaikan program sarjana pada Universitas AMIKOM Yogyakarta. Penyelesaian skripsi ini juga tidak lepas dari bantuan berbagai pihak, karena itu pada kesempatan ini penulis ingin menyampaikan rasa hormat dan terima kasih kepada :

1. Prof. Dr. M. Suyanto, MM. selaku Rektor Universitas AMIKOM Yogyakarta.
2. Ibu Krisnawati, S.Si, M.T. selaku Dekan Fakultas Ilmu Komputer Universitas AMIKOM Yogyakarta.
3. Bapak Sudarmawan, M.T. selaku Ketua Program Studi Informatika Universitas AMIKOM Yogyakarta.
4. Ibu Hartatik, S.T., M.Cs. selaku dosen pembimbing yang selalu bijaksana dalam memberikan bimbingan serta nasehat dan waktunya selama pengerjaan skripsi ini.
5. Ibu Erni Seniwati, S.Kom, M.Cs. dan Ibu Rumini, M.Kom. selaku dosen penguji. Terimakasih atas segala saran yang diberikan selama pengujian untuk memperbaiki penelitian menjadi lebih baik lagi.

Penulis menyadari bahwa skripsi ini masih ada kekurangan. Maka, penulis menerima segala kritik dan saran yang membangun serta teguran dari semua pihak. Semoga skripsi ini bisa bermanfaat baik bagi penulis serta pembaca. Atas saran dan kritik penulis ucapkan terima kasih.

Yogyakarta, 3 Maret 2020

Fariz Aditama

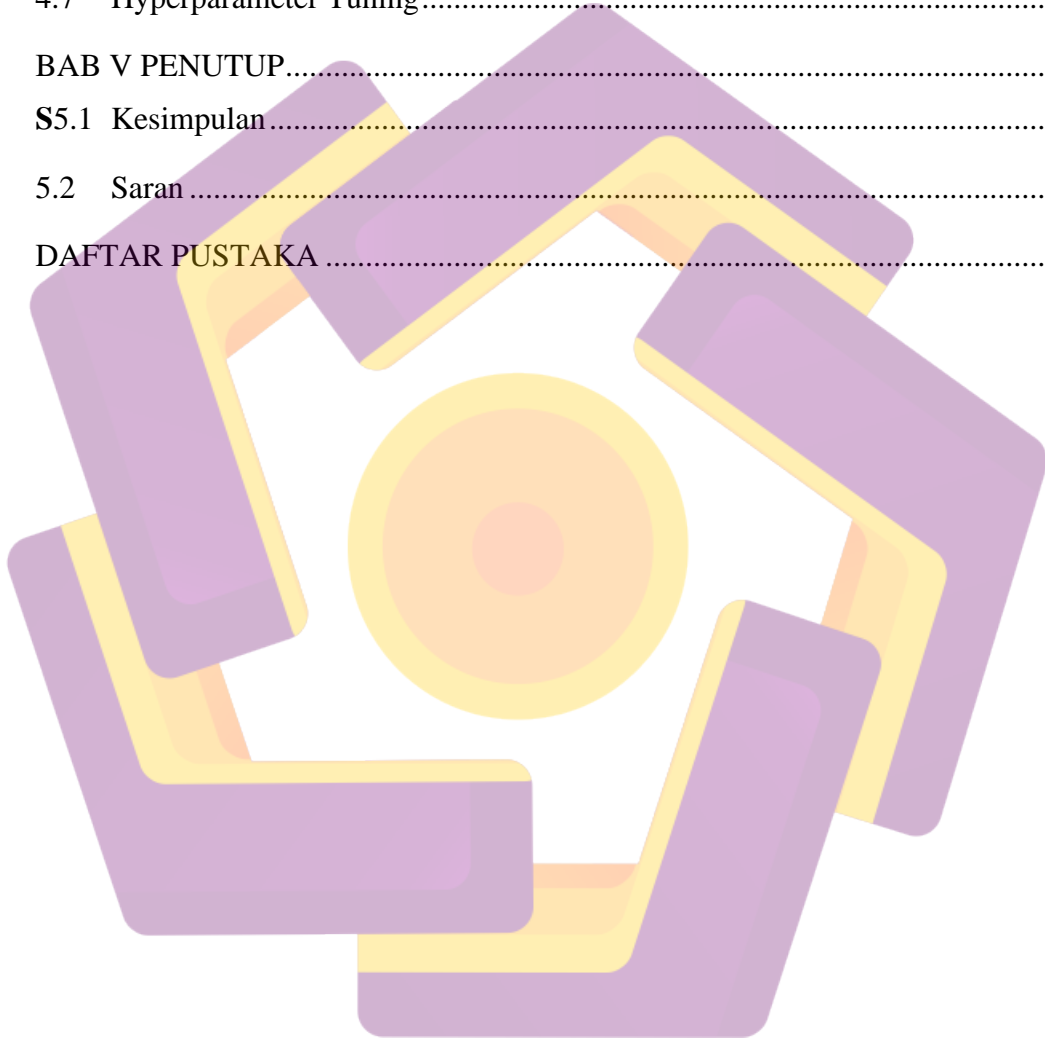


## DAFTAR ISI

JUDUL .....	i
LEMBAR PERSETUJUAN.....	ii
LEMBAR PENGESAHAN .....	iii
PERNYATAAN.....	iv
MOTTO .....	v
PERSEMBAHAN .....	vi
KATA PENGANTAR .....	vii
DAFTAR ISI.....	viii
DAFTAR TABEL.....	xi
DAFTAR GAMBAR .....	xii
INTISARI.....	xiii
<i>ABSTRACT</i> .....	xiv
<b>BAB I PENDAHULUAN</b> .....	<b>1</b>
1.1 Latar Belakang.....	1
1.2 Rumusan Masalah.....	2
1.3 Batasan Masalah .....	3
1.4 Maksud dan Tujuan Penelitian .....	3
1.5 Manfaat Penelitian .....	3
1.6 Metode Penelitian .....	3
1.6.1 Pengumpulan Data .....	3
1.6.2 Preprocessing .....	4
1.6.3 Analisis Data .....	4
1.6.4 Pengujian Algoritma .....	4
1.7 Sistematika Penulisan .....	4
<b>BAB II METODE PENELITIAN</b> .....	<b>6</b>
2.1 Tinjauan Pustaka.....	6

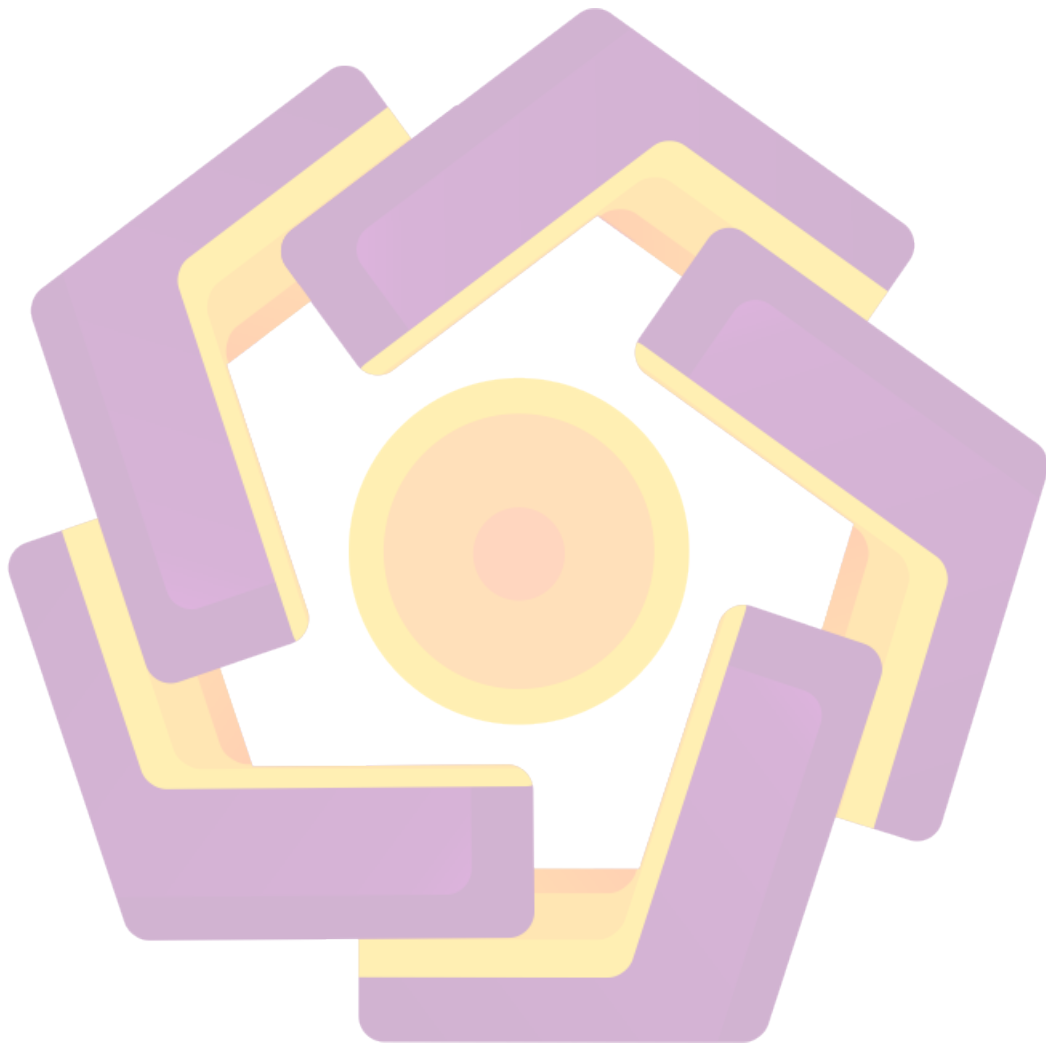
2.2	Dasar Teori .....	9
2.2.1	Big Data.....	9
2.2.2	Data Mining dalam Big Data .....	10
2.2.3	Sistem Rekomendasi .....	12
2.2.4	Collaborative Filtering .....	13
2.2.5	Memory-Based .....	15
2.2.6	Model-based Filtering .....	16
2.2.7	Matrix Factorization .....	17
2.2.8	Singular Value Decomposition .....	17
2.2.9	Non-Negative Matrix Factorization .....	19
2.2.10	Alternating Least Square .....	20
2.2.11	Uji Akurasi .....	21
<b>BAB III METODE PENELITIAN.....</b>		<b>22</b>
3.1	Tahapan Penelitian.....	22
3.2	Alat Penelitian .....	22
3.3	Pengumpulan Data.....	23
3.4	Preprocessing Data .....	23
3.5	Analisis Data.....	24
3.6	Data Training dan Data Test.....	24
3.7	Implementasi Algoritma Matrix Factorization .....	25
3.8	Hasil Evaluasi .....	25
3.9	Tuning Parameter.....	25
<b>BAB IV HASIL DAN PEMBAHASAN .....</b>		<b>27</b>
4.1	Dataset .....	27
4.2	Pengumpulan Data.....	27

4.3	Preprocessing Data .....	28
4.4	Analisis Data.....	29
4.5	Implementasi dan Pengujian Algoritma .....	30
4.6	Pengaruh Parameter Optimisasi .....	34
4.7	Hyperparameter Tuning.....	35
<b>BAB V PENUTUP.....</b>		<b>39</b>
S5.1	Kesimpulan.....	39
5.2	Saran .....	39
<b>DAFTAR PUSTAKA .....</b>		<b>41</b>



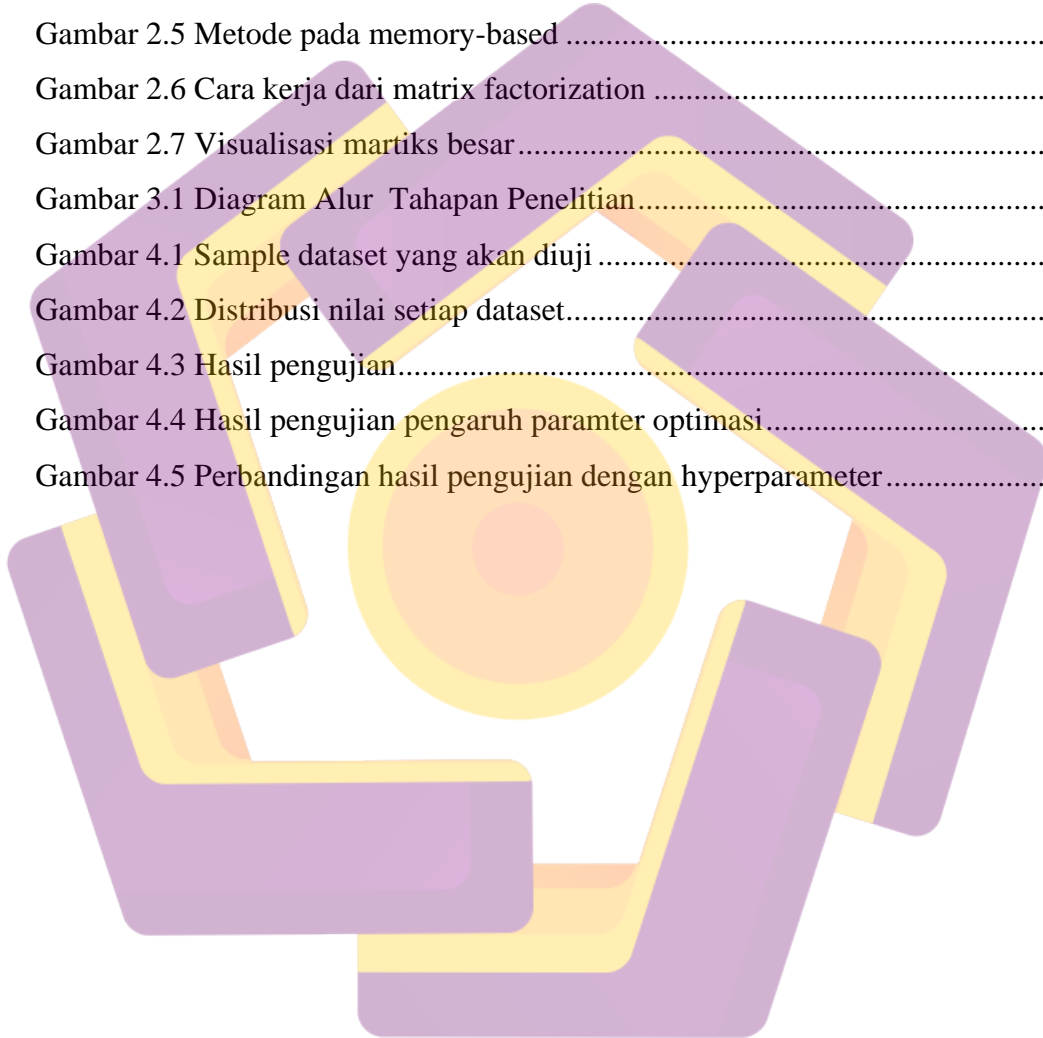
**DAFTAR TABEL**

Tabel 4.1 Analisis dataset .....	29
Tabel 4.2 Pseudocode Algoritma SVD .....	31
Tabel 4.3 Pseudocode Hyperparameter GridSearchCV .....	36



## DAFTAR GAMBAR

Gambar 2.1 Ilustrasi Data Mining.....	10
Gambar 2.2 P Proses Knowledge Discovery Process .....	11
Gambar 2.3 Teknik sistem rekomendasi.....	13
Gambar 2.4 Kategori dari <i>Collaborative Filtering</i> .....	15
Gambar 2.5 Metode pada memory-based .....	16
Gambar 2.6 Cara kerja dari matrix factorization .....	17
Gambar 2.7 Visualisasi matriks besar .....	20
Gambar 3.1 Diagram Alur Tahapan Penelitian.....	22
Gambar 4.1 Sample dataset yang akan diuji .....	28
Gambar 4.2 Distribusi nilai setiap dataset.....	30
Gambar 4.3 Hasil pengujian.....	32
Gambar 4.4 Hasil pengujian pengaruh paramter optimasi.....	35
Gambar 4.5 Perbandingan hasil pengujian dengan hyperparameter .....	37



## INTISARI

Sistem rekomendasi saat ini sudah banyak digunakan pada aplikasi *online* seperti pada *e-commerce*, aplikasi pemesanan makanan online, sosial media, dan masih banyak lagi. Semakin berkembangnya internet sistem rekomendasi menjadi fasilitas yang cukup berpengaruh dalam kepuasan pelanggan. Salah satu teknik yang terpopuler saat ini adalah *Collaborative Filtering* (CB) yang tidak memerlukan informasi tambahan tentang item dan penggunanya. Salah satu masalah terbesar CB adalah mengatasi data yang memiliki nilai jarang atau sparsity yang tinggi. *Model-Based* adalah salah satu dari kategori CB yang memiliki kelebihan dalam mengatasi data sparsity. Oleh sebab itu, pada penelitian ini akan menguji dan membandingkan tiga algoritma yang termasuk dalam *model-based* dalam mengatasi data sparsity.

Pada penelitian ini terdapat tiga algoritma yang akan diuji yaitu *Singular Value Decomposition (SVD)*, *Non-Negative Matrix Factorization*, dan *Alternating Least Square* yang akan diterapkan pada data *movielens* dan *bookcrossing*. Setelah melewati pengujian, kemudian setiap algoritma akan diuji kembali dengan *hyperparameter* agar mendapatkan hasil terbaik. Penelitian ini bertujuan untuk mengetahui algoritma yang memiliki nilai akurasi tertinggi.

Pengujian dilakukan dengan membagi data menjadi 5 yaitu 80% data *training* dan 20% data *test*. Hasil yang diperoleh dengan menggunakan *hyperparameter*, SVD secara keseluruhan memiliki hasil terbaik dari 3 dataset yang diuji, kemudian ALS dan NMF memiliki hasil yang cukup berimbang.

**Kata-kunci:** Rekomendasi, Collaborative Filtering, Model-based, SVD, Movielens

## **ABSTRACT**

*Recommender System is currently widely used in online applications such as e-commerce, online food ordering applications, social media, etc. The development of the internet recommendation is a facility that is quite influential in customer satisfaction. One of the most popular techniques is Collaborative Filtering (CB) which does not require additional information about items and their users. One of the biggest problems of CB is dealing with data that has high sparsity value. Model-based is one of the CB categories that has advantages in overcoming sparsity data. Therefore, this study will test and compare three algorithms that are included in model-based in dealing with data sparsity.*

*In this study, there are three algorithms that will be tested, Singular Value Decomposition (SVD), Non-Negative Matrix Factorization (NMF), and Alternating Least Square which will be applied to dataset movielens and bookcrossing. After passing the test each algorithm will be tested again with hyperparameter to get the best result. This study aims to determine which algorithm has the highest accuracy value.*

*Testing was done by dividing the data into 5, 80% for training data, and 20% test data. The result obtained by using hyperparameter, overall SVD has the best results from the 3 tested datasets, then ALS and NMF have fairly balanced result.*

**Keyword:** *Recommender, Collaborative Filtering, Model-based, SVD, Movielens*