A COMPARISON STUDY OF MODEL BASED COLLABORATIVE FILTERING USING ALTERNATING LEAST SQUARES AND SINGULAR VALUE DECOMPOSITION

THESIS

By:

Aulia Nur Fitrianto

17.61.0108

BACHELOR OF INFORMATICS

FACULTY OF COMPUTER SCIENCE

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2021

A COMPARISON STUDY OF MODEL BASED COLLABORATIVE FILTERING USING ALTERNATING LEAST SQUARES AND SINGULAR VALUE DECOMPOSITION

THESIS

to fulfil the requirements for a Bachelor's degree

in the Informatics study program

By:

Aulia Nur Fitrianto

17.61.0108

BACHELOR OF INFORMATICS

FACULTY OF COMPUTER SCIENCE

UNIVERSITAS AMIKOM YOGYAKARTA

YOGYAKARTA

2021

APPROVAL

THESIS

A COMPARISON STUDY OF MODEL BASED COLLABORATIVE FILTERING USING ALTERNATING LEAST SQUARES AND SINGULAR VALUE DECOMPOSITION

prepared and arranged by:

Aulia Nur Fitrianto
17.61.0108

has been approved by undergraduate thesis supervisor
on
April, 8 2021

Supervisor,

Hartatik, ST., M.Cs.
NIK. 190302232

VALIDATION

THESIS

A COMPARISON STUDY OF MODEL BASED COLLABORATIVE FILTERING USING ALTERNATING LEAST SQUARES AND SINGULAR VALUE DECOMPOSITION

prepared and arranged by:

Aulia Nur Fitrianto

17.61.0108

has been maintained by examiners

on 23 April 2021

The Examiners

Examiner	Signature
Mardhiya Hayaty, S.T, M.Kom. **NIK. 190302108**	_____
Sharazita Dyah Anggita, M.Kom. **NIK. 190302285**	_____
Hartatik, S.T, M.Cs. **NIK. 190302232**	_____

This thesis has been accepted as one of
the requirements for obtaining a Bachelor of Computer degree
on 23 April 2021

DEAN OF FACULTY OF COMPUTER SCIENCE

Hanif Al Fatta, M.Kom
NIK. 190302096

DECLARATION

I, the undersigned bellow, states that, this thesis is my own work (ORIGINAL) and the contents of this thesis have never been applied by any other person to receive an academic degree at a certain education institution, and as far as I know, there are no works or thoughts which have been written and/or published by anyone, except those in writing which are listed in this manuscript and which are mentioned in the reference list.

Anything that applies to the manuscripts and works that have been made is my own responsibility.

Yogyakarta, 8 April 2021

Aulia Nur Fitriana

NIM. 17.61.0108

MOTTO

"*Qui Audet Adipiscitur*"

- British Special Air Services -

"You miss 100% of the shots you don't take."

- Wayne Gretzky –

"Live by the compass, not the clock. The direction should be right."

- Prince Ali Benjamin –

" Be grateful, be brave, do the best."

- Alisha Benjamin -

DEDICATION

Praise Allah SWT, for the grace of a great deal of mercy and His goodness, so that the author can complete the study entitled A Comparison Study of Model Based Collaborative Filtering using Alternating Least Squares and Singular Value Decomposition. Appreciation and gratitude give to :

1. The author's parents, who has been raised and educated, as well as provide support and prayers to the author.
2. The author's brothers, who has been supported to the author.
3. The author's close friends Bayu and Adit who have accompanied me in joy and sorrow and always being the author's partner during college.
4. the authors research partner Nisah, Yahya, Visq, Ibam, Nurma, Christy, Arya, Erick, Anang, Fayami, Nur Hidayah, Wahidin and Dimas.
5. CEO of Binar Academy Ms. Alamandas, who has become my game changer.
6. The authors himself.

ACKNOWLEDGEMENTS

Praise and deep gratitude to Allah SWT for the profusion of grace, and guidance of Him given to the writer that made the undergraduate thesis could be complete. Greetings and sholawah might always be devoted to the prophet of Muhammad SAW. The undergraduate thesis entitled A Comparison Study of Model Based Collaborative Filtering using Alternating Least Squares and Singular Value Decomposition, it is designed to fulfill the undergraduate program (S1) in the Study Program of Informatika, Faculty of Computer Science, Universitas AMIKOM Yogyakarta. The author would like to thank you profusely for all the help that has been given, either directly or indirectly, to complete this final undergraduate thesis during the preparation process. In particular gratitude that are due to:

1. Prof. Dr. M. Suyanto, M.M As Rector of Universitas AMIKOM Yogyakarta
2. Mr. Hanif Al Fatta, M.Kom. As Dean of Faculty of Computer Science Universitas AMIKOM Yogyakarta
3. Mrs. Windha Mega Pradnya Dhuhita, M.Kom, As Head of Study program S1 Informatika Universitas AMIKOM Yogyakarta
4. Mrs. Hartatik S.T., M.Cs. As the author thesis preparation who guidance and encouragement in the preparation of this thesis.
5. The entire teaching staff of the Study Program of Informatika Universitas AMIKOM Yogyakarta has guide me with the knowledge.
6. Big family of FORUM ASISTEN Universitas AMIKOM Yogyakarta which give the authors joy during lectures.

7. Family BCI 01 class of 2017 which give the authors happiness for three years
8. All the parties who could not have been listed one by one by the author, who contributed a lot of support to complete this study.

The authors concluded that this work has not been flawless, either in terms of content or presentation. Ideas and positive suggestion are expected at the conclusion of this thesis. Recently, the writers hope that this research will offer valuable things to the reader and, in particular, to the researcher.

Yogyakarta, 8 April 2021

Author

TABLE OF CONTENT

TITLE I

APPROVAL II

VALIDATION III

DECLARATION ..... ERROR! BOOKMARK NOT DEFINED.

MOTTO V

DEDICATION VI

ACKNOWLEDGEMENTS VII

LIST OF TABLES XII

LIST OF FIGURES XIII

ABSTRACT XIV

CHAPTER I 1

INTRODUCTION 1

1.1 Background 1

1.2 Problems 2

1.3 Problems Limitation 3

1.4 Research Purpose 3

1.5 Research Benefits 3

1.6 Research Methodology 3

1.6.1 Literature Review 4

1.6.2 Problem analysis 4

1.6.3 System Design 4

1.6.4 Implementation 4

1.6.5 Testing 4

1.7 Systematics Writing 5

CHAPTER II 6

RELATED WORKS 6

2.1 Literature Review 6

2.2 Recommender System 10

2.2.1 Collaborative Filtering Recommender System 12
2.2.2 Model Based Filtering 14
2.2.3 Utility Matrix 15
2.2.4 Dimensional Reduction 16
2.2.5 Matrix Factorization 17
2.2.6 Alternating Least Squares 18
2.2.7 Singular Value Decomposition 20
2.2.8 Memory Based Filtering 21
2.2.9 User Based Filtering 21
2.2.10 Item Based Filtering 22
2.2.11 Content Based Filtering 22
2.3 Evaluation Measurement 23
CHAPTER III 24
RESEARCH METHODOLOGY 24
3.1 Research Pipeline 24
3.2 Data Collection 24
3.3 Preprocessing 25
3.4 Exploratory Data Analysis 25
3.5 Data Splitting 25
3.6 Implementation 25
3.7 Evaluation 26
3.8 Experiments 26
CHAPTER IV 27
IMPLEMENTATION AND DISCUSSION 27
4.1 System Implementation 27
4.1.1 Hardware and Software Requirement 27
4.2 Datasets 27
4.3 Data Preprocessing 29
4.4 Exploratory Data Analysis 30
4.5 Training 33
4.6 Testing and Evaluation 34
4.7 Experiment and Result 36
4.7.1 Impact of Epoch 36
4.7.2 Impact of Rank 38
4.7.3 Impact of MaxItter 39

4.7.4 Experiment Results *40*

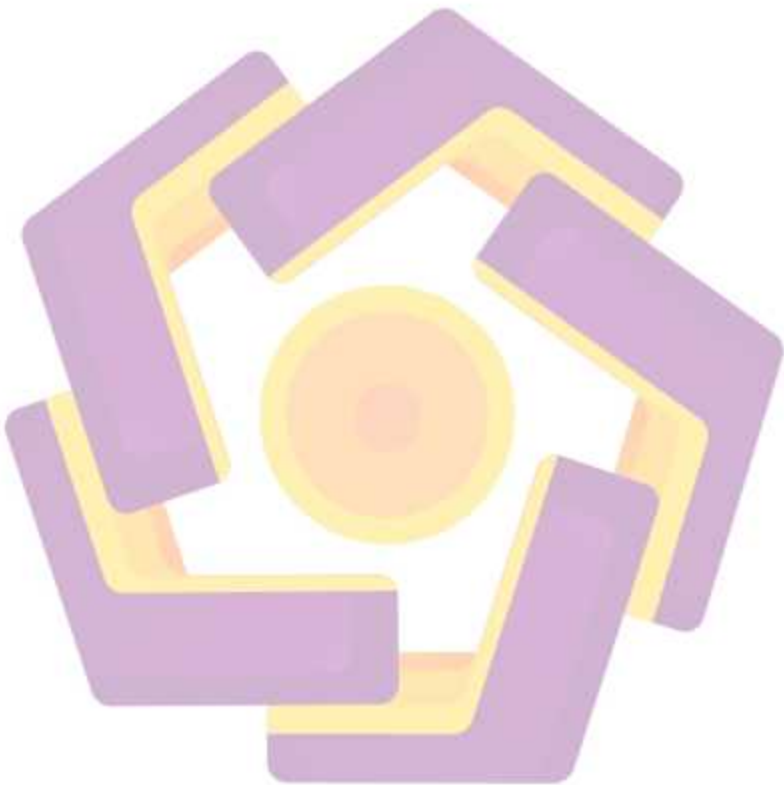
CHAPTER V **42**

CONCLUSION **42**

5.1 Conclusion 42

5.2 Suggestions 43

REFERENCES **44**



LIST OF TABLES

Table 1. 1 Literature Review Table 8
Table 4. 1 Exploratory Data Analysis 31
Table 4. 2 Testing Result with Default Parameter 34
Table 4. 3 Hyperparameter Value 36
Table 4. 4 Experiment Result 40

LIST OF FIGURES

Figure 2. 1 Classification of Recommendation System 11
Figure 2. 2 collaborative filtering schema 13
Figure 2. 3 Collaborative filtering categorized 13
Figure 2. 4 Model Based Collaborative Filtering Taxonomy 15
Figure 2. 5 Utility Matrix Recommendation System Example 15
Figure 2. 6 Dimensional Reduction in SVD 16
Figure 2. 7 Matrix Factorization Diagram 17
Figure 2. 8 Visualize of Dataset in Matrix form 18
Figure 2. 9 Content Based Filtering Schema 22
Figure 3. 1 Research Pipeline 24
Figure 4. 1 Movielens Dataset Sample 28
Figure 4. 2 Jester Joke Dataset Sample 28
Figure 4. 3 Book Crossing Dataset Sample 28
Figure 4. 4 Null value Handling 29
Figure 4. 5 Movielens Utility Matrix 30
Figure 4. 6 Jester Jokes Utility Matrix 30
Figure 4. 7 Book Crossing Utility Matrix 30
Figure 4. 8 Movielens Rating Distribution 32
Figure 4. 9 Jester Jokes Rating Distribution 32
Figure 4. 10 Book Crossing Rating Distribution 32
Figure 4. 11 ALS Training Code 33
Figure 4. 12 SVD Training Code 33
Figure 4. 13 Mean Absolute Error 35
Figure 4. 14 Root Mean Squared Error 35
Figure 4. 15 Impact of Epoch 37
Figure 4. 16 Impact of MaxItter 38
Figure 4. 17 Impact of MaxItter 39

ABSTRACT

Recommender systems are systems that are built to recommend items to users based on a variety of criteria. These systems predict the most likely product that users are likely to buy and are interested in. The most commonly used in recommender system is collaborative filtering.

In this research, the author proposes to determine the quality of recommender system by comparing model based collaborative filtering techniques, i.e Alternating Least Squares and Singular Value Decomposition with three different characteristics dataset.

The result of this study shown Model Based Collaborative Filtering (Alternating Least Squares and Singular Value Decomposition) able to improve the quality of recommender system by considering the hyperparameter tuning result. Overall Alternating Least Squares performs slightly better than Singular Value Decomposition in movielens and jester dataset.

Keywords *: Alternating Least Squares, Singular Value Decomposition, Model Based , Recommender System, Collaborative Filtering*