

**A COMPARISON OF BAT AND FIREFLY ALGORITHM IN  
NEIGHBOURHOOD BASED COLLABORATIVE  
FILTERING**

**THESIS**



By:  
**Bayu Permana Sejati**  
**17.61.0105**

**BACHELOR OF INFORMATICS  
FACULTY OF COMPUTER SCIENCE  
UNIVERSITAS AMIKOM YOGYAKARTA**

**2020**

**A COMPARISON OF BAT AND FIREFLY ALGORITHM IN  
NEIGHBOURHOOD BASED COLLABORATIVE  
FILTERING**

**THESIS**

to fulfil the requirements for a Bachelor's degree  
in the Informatics study program



By:  
**Bayu Permana Sejati**  
**17.01.0105**

**BACHELOR OF INFORMATICS  
FACULTY OF COMPUTER SCIENCE  
UNIVERSITAS AMIKOM YOGYAKARTA  
2020**

**APPROVAL**

**THESIS**

**A COMPARISON OF BAT AND FIREFLY ALGORITHM IN  
NEIGHBOURHOOD BASED COLLABORATIVE  
FILTERING**

prepared and arranged by:

**Bayu Permata Sejati**

**17.61.0105**

has been approved by undergraduate thesis supervisor  
on 8 April 2020

**Supervisor,**

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

**VALIDATION**

**THESIS**

**A COMPARISON OF BAT AND FIREFLY ALGORITHM IN  
NEIGHBOURHOOD BASED COLLABORATIVE**

**FILTERING**

prepared and arranged by:

**Bayu Permana Sejati**

**17.61.0105**

has been maintained by examiners

on 20 October 2020

**The Examiners**

**Examiner**

**Signature**

Ferlan Fauzi Abdullah, M.Kom

**NIK. 190302276**

Uyock Anggoro Saputro, M.Kom

**NIK. 190302419**

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 15 November 2020

**DEAN OF FACULTY OF COMPUTER SCIENCE**

Krisnawati, S.Si, M.T.

**NIK. 190302038**

## 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, 16 November 2020



  
Bayu Permana Sejati  
NIM. 17.61.0105

## MOTTO

“Vivat academia, Vivant professores”

-Franz Liszt-

“Always code as if the guy who ends up maintaining your code will be a violent psychopath who knows where you live.”

- Martin Golding-

“Experience is the name everyone gives to their mistakes.”

- Oscar Wilde-

“From zero to 0,1”

-Bayu Permana Sejati-

## 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 of Bat and Firefly Algorithm in Negighbourhood Based Collaborative Filtering. 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 brother and sister, who has been supported to the author.
3. The author's close friends Aan, Adit, and Ali who have accompanied me in joy and sorrow and always being the author's partner during college.
4. The author's research partner in recommender system Yahya, who guide me to make a good research, and always advising me to keep on track.
5. The authors research partners Nisah, Visq, Ibam, Nurma, and Christy, Arya, Erick, Anang, Fayami, Nur Hidayah, Wahidin, and Dimas.
6. PT. Sasana Solusi Digital which give the author place to learn and develop especially for Mr. Tanim, Bu Els, Pak Oks, Mas Dimas, Mas Asep, Mba Siska, and Mas Ernest.
7. Internship partner team Intan, Fiona, Evan, Luqy, and Dio who being the authors family in Bandung.

## 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 of BAT and Firefly Algorithm in Neighbourhood Based Collaborative Filtering, 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. Mrs. Krisnawati, S.Si, MT. As Dean of Faculty of Computer Science Universitas AMIKOM Yogyakarta.
3. Mr. Sudarmawan, S.T., M.T. As Head of Study program S1 Informatika Universitas AMIKOM Yogyakarta.
4. Mrs. Hartatik, S.T., M.Cs. As the author thesis supervisor 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 Pengurus and Anggota 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 suggestions 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, 16 October 2020



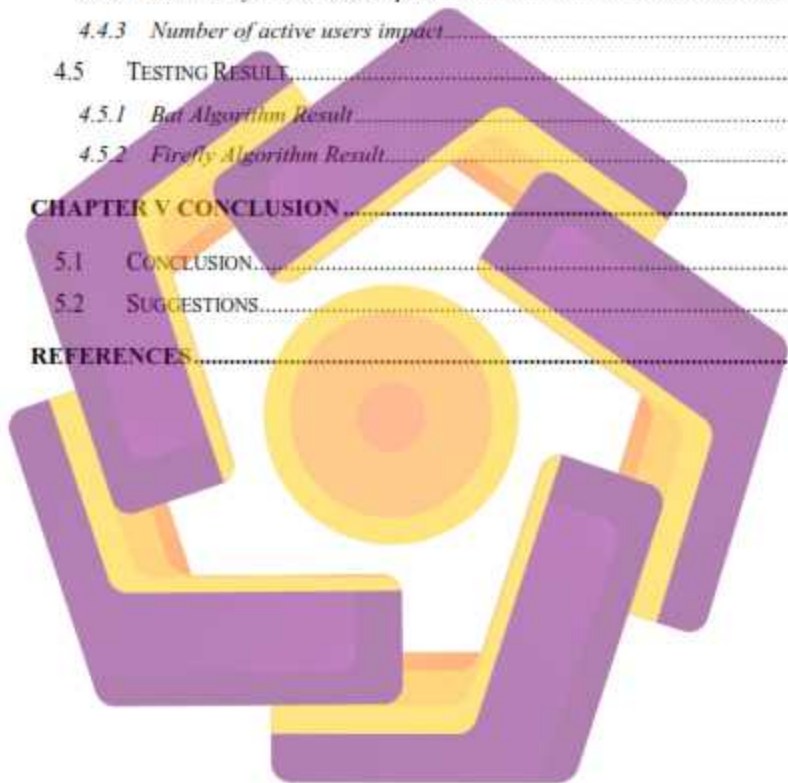
Author

## TABLE OF CONTENT

<b>TITLE</b> .....	<b>I</b>
<b>COVER</b> .....	<b>I</b>
<b>APPROVAL</b> .....	<b>II</b>
<b>VALIDATION</b> .....	<b>III</b>
<b>DECLARATION</b> .....	<b>IV</b>
<b>MOTTO</b> .....	<b>V</b>
<b>DEDICATION</b> .....	<b>VI</b>
<b>ACKNOWLEDGEMENTS</b> .....	<b>VII</b>
<b>LIST OF TABLES</b> .....	<b>XII</b>
<b>LIST OF FIGURES</b> .....	<b>XIII</b>
<b>ABSTRACT</b> .....	<b>XV</b>
<b>CHAPTER I INTRODUCTION</b> .....	<b>1</b>
1.1 BACKGROUND .....	1
1.2 PROBLEMS .....	4
1.3 PROBLEMS LIMITATION .....	4
1.4 RESEARCH PURPOSE .....	5
1.5 RESEARCH BENEFITS .....	5
1.6 RESEARCH METHODOLOGY .....	5
1.6.1 Literature Review .....	5
1.6.2 Problems Analysis .....	5
1.6.3 System Design .....	5
1.6.4 Implementation .....	6
1.6.5 Testing .....	6
1.6.6 Report Formulation .....	6
1.7 SYSTEMATICS WRITING .....	6
<b>CHAPTER II RELATED WORKS</b> .....	<b>8</b>

2.1	LITERATURE REVIEW .....	8
2.2	RECOMMENDER SYSTEMS .....	13
2.2.1	<i>Utility Matrix</i> .....	14
2.2.2	<i>Content Based Recommender System</i> .....	15
2.2.3	<i>Collaborative Filtering Recommender Systems</i> .....	16
2.2.4	<i>Collaborative Filtering Steps</i> .....	20
2.2.5	<i>Pearson Correlation Coefficient</i> .....	21
2.3	SWARM INTELLIGENCE.....	21
2.3.1	<i>BAT Algorithm</i> .....	24
2.4	FIREFLY ALGORITHM.....	27
2.5	PREDICTION CALCULATION.....	29
2.6	EVALUATION MEASUREMENT.....	31
<b>CHAPTER III RESEARCH METHODOLOGY .....</b>		<b>32</b>
3.1	DATASET.....	32
3.2	EXPLORATORY DATA ANALYSIS.....	32
3.2.1	<i>Most popular genre</i> .....	33
3.2.2	<i>The number of movies released each year</i> .....	33
3.2.3	<i>Number of gender distribution</i> .....	34
3.2.4	<i>Number of ratings distribution</i> .....	35
3.4	PRE-PROCESSING.....	37
3.5	TRANSFORMATION UTILITY MATRIX.....	37
3.5	TRANSFORMATION SIMILARITY MATRIX.....	38
3.6	WEIGHTING.....	39
3.6.1	<i>Swarm Intelligence Weighting</i> .....	39
3.9	RESULT ANALYSIS.....	44
<b>CHAPTER IV IMPLEMENTATION AND DISCUSSION .....</b>		<b>45</b>
4.1	SYSTEM IMPLEMENTATION .....	45
4.1.1	<i>Hardware and Software Requirements</i> .....	45
4.1.2	<i>Data Implementation</i> .....	46
4.1.3	<i>Pre-processing</i> .....	47
4.1.4	<i>Transformation Utility Matrix</i> .....	47

4.1.5	Transformation similarity matrix.....	48
4.1.6	Weighting.....	49
4.2	TRAINING.....	51
4.3	TESTING.....	51
4.4	RESULT AND EXPERIMENTS.....	52
4.4.1	Number of <i>K</i> Impact.....	54
4.4.2	Number of Train Data Impact.....	57
4.4.3	Number of active users impact.....	61
4.5	TESTING RESULT.....	64
4.5.1	Bat Algorithm Result.....	65
4.5.2	Firefly Algorithm Result.....	68
<b>CHAPTER V CONCLUSION.....</b>		<b>71</b>
5.1	CONCLUSION.....	71
5.2	SUGGESTIONS.....	72
<b>REFERENCES.....</b>		<b>73</b>



## LIST OF TABLES

Table 2. 1 Comparison table between previous studies.....	11
Table 3.1 Data distribution in MovieLens 100k.....	32
Table 3.2 Data before the transformation process .....	37
Table 3.3 Data after the transformation process .....	38
Table 3.4 Similarity matrix transformation.....	38
Table 4.1 Comparison of parameters in Bat and Firefly Algorithm.....	50
Table 4.2 Weighting results comparison between Bat and Firefly Algorithm .....	50
Table 4.3 Weighting of users comparison between Bat and Firefly Algorithm ...	51
Table 4.4 Experiments result for k=10 and cv=5.....	54
Table 4.5 Experiments result for impact of k=20 .....	54
Table 4. 6 Experiments result for impact of k=30 .....	54
Table 4.7 MAE result for impact of percentage of train data .....	57
Table 4.8 RMSE result for impact of percentage of train data .....	57
Table 4. 9 Execution time for impact of percentage of train data.....	57
Table 4. 10 Result of number active users impact.....	61
Table 4. 11 Result of number non active users impact.....	62
Table 4. 12 Top 10 Recommendation BAT Algorithm with 50 active users .....	65
Table 4. 13 Top 10 Recommendation BAT Algorithm with 100 active users .....	66
Table 4. 14 Top 10 Recommendation BAT Algorithm with 200 active users .....	67
Table 4. 15 Top 10 Recommendation Firefly Algorithm with 50 active users.....	68
Table 4. 16 Top 10 Recommendation Firefly Algorithm with 100 active users... ..	69
Table 4. 17 Top 10 Recommendation Firefly Algorithm with 200 active users... ..	70

## LIST OF FIGURES

Figure 1. 1 Comparison of retail revenue in 1990 and 2021 (Stores; US Securities and Exchange Commission filings; McKinsey analysis).....	2
Figure 2. 1 Utility matrix that represents the ratings of films on a scale of 1 - 5...14	
Figure 2. 2 Schematic of the recommendation system using the content-based method ( <a href="https://www.researchgate.net/">https://www.researchgate.net/</a> ).....	15
Figure 2. 3 Schematic of a recommendation system with a collaborative filtering method ( <a href="https://www.researchgate.net/">https://www.researchgate.net/</a> ).....	17
Figure 2. 4 Centralized and decentralized concept in Swarm Intelligence.....	22
Figure 2. 5 Particle Swarm Optimization concept inspired by a flock of birds [25].	22
Figure 2. 6 Artificial Bee Colony concept inspired by bees [26].....	23
Figure 2. 7 Concept of Bat Algorithm inspired by swarm of bats [27].	23
Figure 2. 8 Example of echolocation in bats for detection of prey. Taken from the website of Ask a Biologist Arizona State University ( <a href="https://askabiologist.asu.edu/echolocation">https://askabiologist.asu.edu/echolocation</a> ).....	24
Figure 3. 1 The number of films from each genre.....	33
Figure 3. 2 The number of movies released each year.....	34
Figure 3. 3 Genders distribution ratio.....	34
Figure 3. 4 Genders and genres distribution ratio.....	35
Figure 3. 5 Number of ratings distribution.....	35
Figure 3. 6 Research pipeline.....	36
Figure 3. 7 Training Process.....	42
Figure 3. 8 Training Process.....	43

Figure 4. 1 Sample ratings data.....	46
Figure 4. 2 Features extraction.....	47
Figure 4. 3 Transformation utility matrix .....	47
Figure 4. 4 Result of transformation similarity matrix .....	48
Figure 4. 5 Transformation similarity matrix.....	48
Figure 4. 6 Result of Transformation similarity matrix.....	49
Figure 4. 7 Initiation of parameters in the Bat Algorithm.....	49
Figure 4. 8 Initiation of parameters in the Firefly Algorithm .....	50
Figure 4. 9 Training data.....	51
Figure 4. 10 Testing data.....	52
Figure 4. 11 Result for MAE .....	52
Figure 4. 12 Result for RMSE .....	53
Figure 4. 13 Comparison of MAE of each K.....	55
Figure 4. 14 Comparison of RMSE of each K.....	55
Figure 4. 15 Comparison of execution time of each K.....	56
Figure 4. 16 Comparison of MAE of train data percentage .....	58
Figure 4. 17 Comparison of RMSE of train data percentage .....	59
Figure 4. 18 Comparison of execution time of train data percentage .....	60
Figure 4. 19 Comparison MAE of user active and not user active .....	62
Figure 4. 20 Comparison RMSE of user active and not user active .....	63
Figure 4. 21 Comparison execution time of user active and not user active .....	63
Figure 4. 22 Prediction with real rating .....	64

## ABSTRACT

A Recommender System is a program for knowledge filtering that allows users to recommend items to users in a customizable manner and offer recommendations about the items that they might like. Recently, researchers have used the heuristic algorithm in the traditional technique of recommender system based on collaborative filtering or content-based filtering. In this research, the author proposes to prescribe the quality of recommender system by comparing swarm intelligence techniques, i.e. Bat and Firefly Algorithm. Swarm Intelligence uses to determine the user weight that provides by Pearson Correlation Coefficient. The result of this study shows Swarm Intelligence (Bat and Firefly Algorithm) able to improve the quality of Recommender System by considering the active users, and the Firefly Algorithm performs slightly better rather than Bat Algorithm by 0,0198 for Mean Absolute Error computation.

*Keywords : Bat Algorithm, Firefly Algorithm, Recommender System, Pearson Correlation Coefficient, Collaborative Filtering.*