



**PROJECT REPORT**  
**IMPLEMENTATION OF K-MEANS ALGORITHM**  
**ELBOW METHOD AND SILHOUETTE COEFFICIENT FOR**  
**RAINFALL CLASSIFICATION**

**DANIEL ADRIAN SETIADY**  
**17.K1.0009**

**Faculty of Computer Science**  
**Soegijapranata Catholic University**  
**2021**

## HALAMAN PENGESAHAN



Judul Tugas Akhir: : IMPLEMENTATION OF K-MEANS ALGORITHM ELBOW METHOD  
AND SILHOUETTE COEFFICIENT FOR RAINFALL CLASSIFICATION

Diajukan oleh : Daniel Adrian Setiady

NIM : 17.K1.0009

Tanggal disetujui : 09 Juli 2021

Telah setuju oleh

Pembimbing : Hironimus Leong S.Kom., M.Kom.

Penguji 1 : Hironimus Leong S.Kom., M.Kom.

Penguji 2 : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

Penguji 3 : Rosita Herawati S.T., M.I.T.

Penguji 4 : Y.b. Dwi Setianto

Penguji 5 : Yonathan Purbo Santosa S.Kom., M.Sc

Penguji 6 : Yulianto Tejo Putranto S.T., M.T.

Ketua Program Studi : Rosita Herawati S.T., M.I.T.

Dekan : R. Setiawan Aji Nugroho S.T., MCompIT., Ph.D

Halaman ini merupakan halaman yang sah dan dapat diverifikasi melalui alamat di bawah ini.

[sintak.unika.ac.id/skripsi/verifikasi/?id=17.K1.0009](http://sintak.unika.ac.id/skripsi/verifikasi/?id=17.K1.0009)

## STATEMENT PAGE OF SCIENTIFIC WORK PUBLICATION FOR ACADEMIC INTEREST

I, the undersigned:

Name : Daniel Adrian Setiady

Department : Informatics Engineering

Fakultas : Computer Science

Jenis Karya : Thesis

[Menyetujui / Tidak Menyetujui] untuk memberikan kepada Universitas Katolik Soegijapranata Semarang Hak Bebas Royalti Noneksklusif atas karya ilmiah yang berjudul “[Judul Tugas Akhir]” beserta perangkat yang ada (jika diperlukan). Dengan Hak Bebas Royalti Noneksklusif ini Universitas Katolik Soegijapranata berhak menyimpan, mengalihkan media/formatkan, mengelola dalam bentuk pangkalan data (database), merawat, dan mempublikasikan tugas akhir ini selama tetap mencantumkan nama saya sebagai penulis / pencipta dan sebagai pemilik Hak Cipta.

Demikian pernyataan ini saya buat dengan sebenarnya.

Semarang, 9<sup>th</sup> July 2021

Yang menyatakan



[Daniel Adrian Setiady]

## DECLARATION OF AUTHORSHIP

I, the undersigned:

Name : DANIEL ADRIAN SETIADY

ID : 17.K1.0009

declare that this work, titled "IMPLEMENTATION OF K-MEANS ALGORITHM ELBOW METHOD AND SILHOUETTE COEFFICIENT FOR RAINFALL CLASSIFICATION", and the work presented in it is my own. I confirm that:

1. This work was done wholly or mainly while in candidature for a research degree at Soegijapranata Catholic University
2. Where any part of this thesis has previously been submitted for a degree or any other qualification at this University or any other institution, this has been clearly stated.
3. Where I have consulted the published work of others, this is always clearly attributed.
4. Where I have quoted from the work of others, the source is always given.
5. Except for such quotations, this work is entirely my own work.
6. I have acknowledged all main sources of help.
7. Where the work is based on work done by myself jointly with others, I have made clear exactly what was done by others and what I have contributed myself.

Semarang, July, 09, 2021



DANIEL ADRIAN SETIADY

17.K1.0009

## ABSTRACT

*Rain is one of the hydrological cycles which is a cycle of water rotation from the earth to the atmosphere and back to the earth continuously. High Rainfall may cause some areas that are in lowlands or those with low water infiltration systems will be very susceptible to flooding. For that it is necessary to have a system to classify weather data and rainfall in each city and district so the city that has high rainfall and extreme weather can be given special attention to prevent any natural disaster like flooding.*

*The collected data will be processed with K-Means algorithm to classify the cities or district that have low, medium, high, or very high rainfall data. In the K-Means algorithm the amount of k or cluster usually determined by randomly, on this project will be used a method that is Elbow Method to determine the value of k or cluster and Silhouette Coefficient Method will be used for testing the quality amount of a cluster. The data that will be used is the rainfall data from dataonline.bmkg.go.id at a certain period of time to be classified using the K-Means algorithm.*

*The elbow method and the silhouette method can be used in selecting a good optimal number of clusters, and both method mostly have the same results in determining the optimal number of clusters, it can be seen that the calculation of accuracy between using the optimal number of clusters is higher rather than not using the amount optimal number of cluster. This can be seen in the results of the clustering in Semarang on February 1 - 28, 2021, when using the amount of  $K = 4$  produce the accuracy result 92.8571429 %, while when using the optimal number of cluster  $K=3$  the accuracy result is higher (97.6190476 %). In the Cilacap city classification on April 1-30 2021, the elbow method and the silhouette coefficient method produce different optimal cluster results, but the accuracy obtained when using the optimal number of clusters from the silhouette coefficient (85.7142857 %) is higher than using the optimal cluster from the elbow method.(74.6031746 %), but when the data is processed with centroid on table 5.10, the elbow method and silhouette coefficient method produce the same amount of optimal number of clusters is 2. This shows that differences in the use of the initial centroid point can affect the results of the elbow method and the silhouette coefficient method*

*Keyword: Classification, K-Means, Elbow Method, Silhouette Coefficient, Weather*

## ACKNOWLEDGEMENT

I want to thank God Almighty, because with His blessings and goodness this project can be completed, I also want to thank my supervisor Hironimus Leong S.Kom., M.Kom. who have been willing to help and guide in this project. I dedicate this thesis to my parents and people around who need it



# TABLE OF CONTENTS

<b>COVER.....</b>	<b>i</b>
<b>APPROVAL AND RATIFICATION PAGE (Heading plain) .....</b>	<b>ii</b>
<b>STATEMENT PAGE OF SCIENTIFIC WORK PUBLICATION FOR ACADEMIC INTEREST .....</b>	<b>iii</b>
<b>DECLARATION OF AUTHORSHIP .....</b>	<b>iv</b>
<b>ABSTRACT.....</b>	<b>v</b>
<b>ACKNOWLEDGEMENT.....</b>	<b>vi</b>
<b>TABLE OF CONTENTS .....</b>	<b>vii</b>
<b>LIST OF FIGURE .....</b>	<b>ix</b>
<b>LIST OF TABLE .....</b>	<b>x</b>
<b>CHAPTER 1 INTRODUCTION.....</b>	<b>1</b>
1.1 Background.....	1
1.2 Problem Formulation.....	1
1.3 Scope.....	2
1.4 Objective.....	2
<b>CHAPTER 2 LITERATURE STUDY.....</b>	<b>3</b>
<b>CHAPTER 3 RESEARCH METHODOLOGY.....</b>	<b>6</b>
3.1 Journal.....	6
3.2 Data.....	6
3.3 Structure Design.....	6
3.4 Program.....	7
3.5 Implementation and analysis.....	7
3.6 Conclusion and writing report.....	7
<b>CHAPTER 4 ANALYSIS AND DESIGN.....</b>	<b>8</b>

4.1	Analysis.....	8
4.2	Design .....	35
<b>CHAPTER 5 IMPLEMENTATION AND TESTING .....</b>		<b>36</b>
5.1	Implementation .....	36
5.2	Testing.....	49
<b>CHAPTER 6 CONCLUSION.....</b>		<b>87</b>
6.1	Conclusion .....	87
6.2	Suggestion.....	87
<b>REFERENCES.....</b>		<b>a</b>
<b>APPENDIX.....</b>		<b>c</b>





## LIST OF FIGURE

<b>Figure 1.1.</b> Elbow Method .....	26
<b>Figure 1.2.</b> Flowchart Process .....	35
<b>Figure 1.3.</b> Input Form .....	36
<b>Figure 1.4.</b> WCSS result .....	45
<b>Figure 1.5.</b> Silhouette Score Result.....	48
<b>Figure 1.6.</b> Input Form .....	49
<b>Figure 1.7.</b> The Result from C1 on K=4 .....	50
<b>Figure 1.8.</b> The result from C2 on K=4 .....	51
<b>Figure 1.9.</b> The result from C3 on K=4 .....	52
<b>Figure 1.10.</b> The result from C=4 on K=4 .....	53
<b>Figure 1.11.</b> Elbow method result for Semarang 1-28 February 2021 .....	54
<b>Figure 1.12.</b> The graph of Silhouette score for Semarang on 1-28 February 2021 .....	55
<b>Figure 1.13.</b> The result of C=1 at K=3 .....	56
<b>Figure 1.14.</b> The result of C=2 at K=3 .....	57
<b>Figure 1.15.</b> The Result of C=3 at K=3 .....	58
<b>Figure 1.16.</b> The elbow method on Cilacap 1-30 April 2021 .....	61
<b>Figure 1.17.</b> Silhouette Coefficient Graph on Cilacap 1-30 April 2021 .....	61
<b>Figure 1.18.</b> Silhouette Coefficient Result on Cilacap 1-30 April when using table 1.42 Centroid.....	65
<b>Figure 1.19.</b> Elbow Method on Semarang 1 January 2021 - 31 May 2021 .....	68
<b>Figure 1.20.</b> Silhouette Coefficient on Semarang 1 January 2021 - 31 May 2021 .....	68

## LIST OF TABLE

<b>Table 1.1</b> : Raw Data Attributes.....	8
<b>Table 1.2</b> : Used Data Attributes.....	9
<b>Table 1.3</b> : Data.....	9
<b>Table 1.4</b> : Initial Centroid.....	10
<b>Table 1.5</b> : 1 <sup>st</sup> Iteration.....	13
<b>Table 1.6</b> : 1 <sup>st</sup> Iteration Closest Cluster Result.....	14
<b>Table 1.7</b> : Final Centroid 1 <sup>st</sup> Iteration.....	15
<b>Table 1.8</b> : 2 <sup>nd</sup> Iteration.....	16
<b>Table 1.9</b> : 2 <sup>nd</sup> Iteration Closest Cluster Result.....	17
<b>Table 1.10</b> : Final Centroid 2 <sup>nd</sup> Iteration.....	17
<b>Table 1.11</b> : Final Centroid 3 <sup>rd</sup> Iteration.....	18
<b>Table 1.12</b> : Final Centroid 4 <sup>th</sup> Iteration.....	18
<b>Table 1.13</b> : Final Centroid 5 <sup>th</sup> Iteration.....	19
<b>Table 1.14</b> : Final Centroid 6 <sup>th</sup> Iteration.....	19
<b>Table 1.15</b> : Final Result of processing data.....	20
<b>Table 1.16</b> : Final Centroid on K=1.....	22
<b>Table 1.17</b> : The amount of WCSS on K=1.....	22
<b>Table 1.18</b> : Final Centroid on K=3.....	23
<b>Table 1.19</b> : The amount of WCSS on K=3.....	23
<b>Table 1.20</b> : The amount of WCSS on K=3.....	24
<b>Table 1.21</b> : Final Centroid on K=4.....	25
<b>Table 1.22</b> : The amount of WCSS on K=4.....	25

<b>Table 1.23</b> : Cluster result for K=2 .....	28
<b>Table 1.24</b> : a(i) result for K=2 .....	29
<b>Table 1.25</b> : The distance from C2 to all other clusters on K=2 .....	29
<b>Table 1.26</b> : b(i) result on k=2.....	30
<b>Table 1.27</b> : S(i) result on K=2 .....	31
<b>Table 1.28</b> : s(i) result on K=3 .....	32
<b>Table 1.29</b> : s(i) result on K=4 .....	33
<b>Table 1.30</b> : Cluster quality result with Silhouette Score .....	33
<b>Table 1.31</b> : The amount of initial Centroids .....	49
<b>Table 1.32</b> : Conclusion Result on K=4.....	53
<b>Table 1.33</b> : WCSS result on Semarang 1-28 February 2021.....	54
<b>Table 1.34</b> : Silhouette Score on Semarang 1-28 February 2021 .....	55
<b>Table 1.35</b> : Conclusion result on semarang from 1-28 February 2021 .....	58
<b>Table 1.36</b> : Accuracy Result from K=3 on Semarang 1-28 February 2021 .....	59
<b>Table 1.37</b> : Accuracy Result from K=4 on Semarang 1-28 February 2021 .....	60
<b>Table 1.38</b> : Conclusion result on Cilacap from 1-30 April 2021 (K=2).....	62
<b>Table 1.39</b> : Conclusion result on Cilacap from 1-30 April 2021 (K=3).....	62
<b>Table 1.40</b> : Accuracy result from K=2 on Cilacap 1-30 April 2021.....	62
<b>Table 1.41</b> : Accuracy result from K=3 on Cilacap 1-30 April 2021 .....	63
<b>Table 1.42</b> : Starting Centroid.....	64
<b>Table 1.43</b> : Elbow Method Result on Cilacap 1-30 April when using table 5.10 Centroid .....	64
<b>Table 1.44</b> : Conclusion result on Cilacap from 1-30 April 2021 (K=2) when using table 1.42 Centroids.....	66

<b>Table 1.45</b> : Accuracy result Cilacap 1-30 April 2021 when using table 1.42 Centroids	66
<b>Table 1.46</b> : Conclusion result on Cilacap from 1-30 April 2021 (K=3) when using table 1.42 Centroids	66
<b>Table 1.47</b> : Accuracy result Cilacap 1-30 April 2021 (K=3) when using table 5.10 Centroids	67
<b>Table 1.48</b> : Result on Semarang 1 January 2021- 31 May 2021 using the optimal number of Kluster = 2	69
<b>Table 1.49</b> : Accuracy result from K=2 on Semarang 1 January -31 May 2021	69
<b>Table 1.50</b> : Result on Semarang 1 January 2021- 31 May 2021 when Kluster = 3	69
<b>Table 1.51</b> : Accuracy result from K=3 on Semarang 1 January - 31 May 2021	70
<b>Table 1.52</b> : Result on Semarang 1 January 2021- 31 May 2021 when Kluster = 4	70
<b>Table 1.53</b> : Accuracy result from K=4 on Semarang 1 January - 31 May 2021	70
<b>Table 1.54</b> : Elbow Method on Tegal 1 January 2021 - 31 May 2021	71
<b>Table 1.55</b> : Silhouette Coefficient on Tegal 1 January 2021 - 31 May 2021	71
<b>Table 1.56</b> : Result on Tegal 1 January 2021- 31 May 2021 using the optimal number of Kluster 2	72
<b>Table 1.57</b> : Accuracy result from K=2 on Tegal 1 January -31 May 2021	72
<b>Table 1.58</b> : Result on Tegal 1 January 2021- 31 May 2021 using Kluster = 3	72
<b>Table 1.59</b> : Accuracy result from K=3 on Tegal 1 January -31 May 2021	73
<b>Table 1.60</b> : Result on Tegal 1 January 2021- 31 May 2021 using Kluster = 4	73
<b>Table 1.61</b> : Accuracy result from K=4 on Tegal 1 January -31 May 2021	74
<b>Table 1.62</b> : Silhouette Coefficient on Cilacap 1 January 2021 - 31 May 2021	75
<b>Table 1.63</b> : Elbow Method on Cilacap 1 January 2021 - 31 May 2021	75
<b>Table 1.64</b> : Result on Cilacap 1 January 2021- 31 May 2021 using the optimal number of Kluster = 2	76

<b>Table 1.65</b> : Accuracy result from K=2 on Cilacap 1 January -31 May 2021 .....	76
<b>Table 1.66</b> : Elbow Method on Jakarta 1 January 2021 - 31 May 2021 .....	77
<b>Table 1.67</b> : Silhouette Coefficient on Jakarta 1 January 2021 - 31 May 2021.....	77
<b>Table 1.68</b> : Result on Jakarta 1 January 2021- 31 May 2021 using optimal Kluster =	278
<b>Table 1.69</b> : Accuracy result from K=2 on Jakarta 1 January -31 May 2021 .....	78
<b>Table 1.70</b> : Result on Jakarta 1 January 2021- 31 May 2021 using Kluster = 3 .....	78
<b>Table 1.71</b> : Accuracy result from K=3 on Jakarta 1 January -31 May 2021 .....	79
<b>Table 1.72</b> : Result on Jakarta 1 January 2021- 31 May 2021 using Kluster = 4 .....	79
<b>Table 1.73</b> : Accuracy result from K=4 on Jakarta 1 January -31 May 2021 .....	80
<b>Table 1.74</b> : Elbow Method on Bandung 1 January 2021 - 31 May 2021 .....	80
<b>Table 1.75</b> : Silhouette Coefficient on Bandung 1 January 2021 - 31 May 2021.....	81
<b>Table 1.76</b> : Result on Bandung 1 January 2021- 31 May 2021 using optimal Kluster = 2	81
.....	
<b>Table 1.77</b> : Accuracy result from K=2 on Bandung 1 January -31 May 2021.....	81
<b>Table 1.78</b> : Result on Bandung 1 January 2021- 31 May 2021 using Kluster = 3 .....	82
<b>Table 1.79</b> : Accuracy result from K=3 on Bandung 1 January -31 May 2021.....	82
<b>Table 1.80</b> : Result on Bandung 1 January 2021- 31 May 2021 using Kluster = 4.....	82
<b>Table 1.81</b> : Accuracy result from K=4 on Bandung 1 January -31 May 2021.....	83
<b>Table 1.82</b> : Final Result .....	85