



UNDERGRADUATE THESIS

ANALYSIS OF THE COMBINATION OF K NEAREST NEIGHBOR (KNN) AND K-MEANS IN THE CLASSIFICATION OF RICE LEAF DISEASES USING AN IMAGE SEGMENTATION METHOD

RIZAL AFANDI

NPM 19081010146

SUPERVISOR

Dr. Ir. Kartini, S.Kom. MT.

Retno Mumpuni, S.Kom., M.Sc

**MINISTRY OF HIGHER EDUCATION, SCIENCE, AND TECHNOLOGY
UNIVERSITAS PEMBANGUNAN NASIONAL VETERAN JAWA TIMUR
FACULTY OF COMPUTER SCIENCE
INFORMATICS STUDY PROGRAM
SURABAYA
2026**

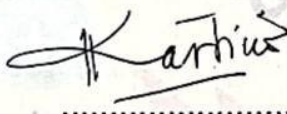
APPROVAL SHEET

ANALYSIS OF THE COMBINATION OF K-NEAREST NEIGHBOR (KNN) AND K-MEANS IN THE CLASSIFICATION OF RICE LEAF DISEASES USING AN IMAGE SEGMENTATION METHOD

By:
RIZAL AFANDI
NPM. 190810010146

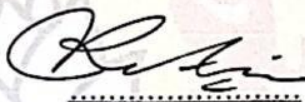
Has been defended before and accepted by the Thesis Examination Committee of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on June 22, 2026:

Dr. Ir. Kartini, S.Kom. MT.
NIP. 19611110 199103 2 001



(Advisor I)

Retno Mumpuni, S.Kom., M.Sc
NIP. 198707162025212045



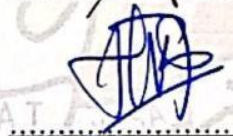
(Advisor II)

Eka Prakarsa Mandvarcha, S.T., M.Kom
NIP. 19800907 202121 1 005



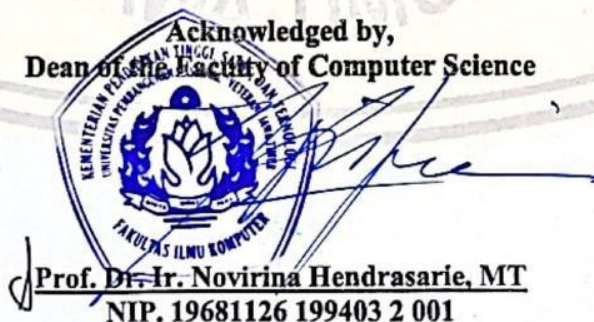
(Head Asesor)

Dr. Intan Yuniar Purbasari, S.Kom.
NIP. 19800602 202521 2 029



(Asesor I)

Acknowledged by,
Dean of the Faculty of Computer Science


Prof. Dr. Ir. Novirina Hendrasarie, MT
NIP. 19681126 199403 2 001

AGREEMENT SHEET

**ANALYSIS OF THE COMBINATION OF K-NEAREST
NEIGHBOR (KNN) AND K-MEANS IN THE
CLASSIFICATION OF RICE LEAF DISEASES USING AN
IMAGE SEGMENTATION METHOD**

By :

**RIZAL AFANDI
NPM. 19081010146**

Approved for

Approved by:

**Coordinator of Informatics Study Program
Faculty of Computer Science**



**Dr. Intan Yuniar Purbasari, S.Kom.
NIP. 19800602 202521 2 029**

STATEMENT OF ORIGINALITY

The undersigned below:

Student Name/NPM : RIZAL AFANDI/190810010146
Study Program : Informatics
Supervisors : 1. Dr. Ir. Kartini, S.Kom. MT.
2. Retno Mumpuni, S.Kom., M.Sc

hereby declares that the content, partially or wholly, of the thesis with the title:

**“ANALYSIS OF THE COMBINATION OF K-NEAREST NEIGHBOR
(KNN) AND K-MEANS IN THE CLASSIFICATION OF RICE LEAF
DISEASES USING AN IMAGE SEGMENTATION METHOD”**

is the result of my own work, is original, and is written following scientific writing principles. If in the future there are discrepancies with this statement, I am willing to accept sanctions in accordance with the applicable regulations at the University of Pembangunan Nasional Veteran East Java.



Surabaya, 22 June 2026

The undersigned,



RIZAL AFANDI

NPM. 19081010146

ABSTRACT

Student Name / NPM : Rizal Afandi/190810010146
Thesis Title : Analysis of the Combination of K-Nearest Neighbor (KNN) and K-Means in Rice Leaf Disease Classification Using Image Segmentation Methods
Advisor : 1. Dr. Ir. Kartini, S.Kom. MT.
2. Retno Mumpuni, S.Kom., M.Sc

ABSTRACT

Rice is a primary food commodity in Indonesia that plays an essential role in national food security. However, rice productivity is often reduced due to leaf diseases such as Bacterial Blight, Blast, Brown Spot, and Tungro. Manual disease identification is considered inefficient as it requires expert knowledge and considerable time. Therefore, an automated system based on image processing and machine learning is needed to assist in the fast and accurate classification of rice leaf diseases.

This study aims to implement a combination of K-Means and K-Nearest Neighbor (KNN) algorithms for classifying rice leaf diseases using image segmentation methods. The dataset used consists of 5,932 rice leaf images divided into four disease classes: Bacterial Blight, Blast, Brown Spot, and Tungro. The research stages include image preprocessing (grayscale conversion, resizing, and flattening), segmentation using K-Means, and classification using KNN. Testing is conducted using various training and testing data ratios as well as different K values to determine the optimal performance.

The results show that the combination of K-Means and KNN is capable of classifying rice leaf diseases with high accuracy. The best performance is achieved using a 90:10 training-to-testing ratio with an optimal K value, producing highly accurate predictions based on confusion matrix evaluation. Therefore, the proposed method is effective in supporting automatic identification of rice leaf diseases.

This research is expected to contribute to the development of image-based plant disease detection systems and assist farmers in diagnosing diseases more quickly and accurately, thereby improving agricultural productivity.

Keywords: K-Means, K-Nearest Neighbor, classification, rice leaf disease, image processing.

PREFACE

All praise be to Allah SWT for all His grace, guidance, and blessings, so that the author can complete this thesis entitled "**Analysis of the Combination of K-Nearest Neighbor (KNN) and K-Means in Rice Leaf Disease Classification Using Image Segmentation Method**" well and on time.

This thesis is submitted as one of the requirements to obtain a Bachelor's degree in the Informatics Engineering Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur. In the process of writing this thesis, the author realizes that many parties have provided assistance, support, and guidance, either directly or indirectly.

Therefore, on this occasion, the author wishes to express gratitude to:

1. Allah SWT who has given health and ease in completing this thesis.
2. Dr. Ir. Kartini, S.Kom., M.T. as the first supervisor who has provided guidance, direction, and motivation in the preparation of this thesis.
3. Retno Mumpuni, S.Kom., M.Sc. as the second supervisor who has provided suggestions, input, and support during the research process.
4. All lecturers of the Informatics Engineering Study Program who have imparted knowledge during the course of study.
5. My beloved mother, who always supports me in completing my research until writing my thesis report. Thank you for your patience in waiting for me to finish this study and I apologize for the delay in completing it. She has always been my shoulder to lean on when I feel tired, always present with her best prayers, a source of unfailing strength, and is the reason why I can finish this thesis immediately. May Allah SWT always bestow good health and a long life, so that the author can even repay the kindness and affection that you have given me all this time. I express my deepest gratitude and I love you very much.
6. My dearest Father, who taught to be someone who was responsible, worked hard, and did not give up easily. His advice and examples are an important foundation for the author in undergoing each process with determination. May Allah SWT always grant me good health and a long life, so that I can

repay you all for your kindness and love. I express my deepest gratitude and love you all so much.

7. To my little sister, thank you so much for being one of the biggest reasons and my main source of motivation to finish this thesis as soon as possible. I hope this will inspire you to go even further and achieve your dreams in the future.
8. Friends and all parties who cannot be mentioned one by one who have helped in the preparation of this thesis.

The author realizes that there are still many shortcomings in the preparation of this thesis. Therefore, the author expects constructive criticism and suggestions for the improvement of this work in the future.

Finally, the author hopes that this thesis can provide benefits for readers and can serve as a reference for future research, especially in the field of digital image processing and plant disease classification.

Surabaya, May 25 2026

Rizal Afandi

TABLE OF CONTENTS

TITLE PAGE.....	i
APPROVAL SHEET	ii
AGREEMENT SHEET	iii
STATEMENT OF ORIGINALITY	iv
ABSTRACT.....	v
PREFACE	vi
TABLE OF CONTENTS.....	viii
LIST OF FIGURES.....	xi
LIST OF TABLES.....	xiv
LIST OF NOTATIONS	xv
CHAPTER I INTRODUCTION	1
1.1. Background.....	1
1.2. Problem Formulation	4
1.3. Research Objectives	5
1.4. Research Benefits	5
CHAPTER II LITERATURE REVIEW	6
2.1. Previous Research	6
2.2. Theoretical Basis.....	8
2.2.1. Rice.....	8
2.2.2. Artificial Intelligence.....	10
2.2.3. K-Nearest Neighbor (KNN)	12
2.2.4. K-Means Clustering.....	14
2.2.5. Digital Image Processing.....	15
2.2.6. Image Acquisition.....	17
2.2.7. Rescaling	17
2.2.8. Resizing	18
2.2.9. Image Segmentation	19
2.2.10. Grayscale	20
CHAPTER III SYSTEM DESIGN AND IMPLEMENTATION	22
3.1. Research Method	22

3.2. System Design	24
3.2.1. System Planning	24
3.2.2. Image Pre-processing	25
3.2.3. Image Segmentation Using K-Means	27
3.2.4. Classification Using K-Nearest Neighbor (KNN).....	28
3.2.5. Implementation	29
3.2.6. Testing and Analysis	29
CHAPTER IV TESTING AND ANALYSIS.....	30
4.1. Testing Method.....	30
4.1.1. Dataset Preparation.....	30
4.1.2. Data Preprocessing	30
4.1.3. Data Processing	33
4.2. Test Results	41
4.2.1. Visualization K-Means Clustering 90 : 10	42
4.2.2. Scenario 1	44
4.2.3. Scenario 2	46
4.2.4. Scenario 3	47
4.2.5. Scenario 4	49
4.2.6. Scenario 5	50
4.2.7. Scenario 6	52
4.2.8. Scenario 7	53
4.2.9. Scenario 8	55
4.2.10. Scenario 9	56
4.2.11. Scenario 10	58
4.2.12. Visualization K-Means 80:20	59
4.2.13. Scenario 11	61
4.2.14. Scenario 12	63
4.2.15. Scenario 13	64
4.2.16. Scenario 14	66
4.2.17. Scenario 15	67
4.2.18. Scenario 16	69
4.2.19. Scenario 17	70

4.2.20. Scenario 18	72
4.2.21. Scenario 19	74
4.2.22. Scenario 20	76
4.2.23. Visualization K-Means 70:30	77
4.2.24. Scenario 21	79
4.2.25. Scenario 22	81
4.2.26. Scenario 23	82
4.2.27. Scenario 24	84
4.2.28. Scenario 25	85
4.2.29. Scenario 26	87
4.2.30. Scenario 27	88
4.2.31. Scenario 28	90
4.2.32. Scenario 29	91
4.2.33. Scenario 30	93
4.2.34. Visualization K-Means 60:40	94
4.2.35. Scenario 31	96
4.2.36. Scenario 32	98
4.2.37. Scenario 33	99
4.2.38. Scenario 34	101
4.2.39. Scenario 35	102
4.2.40. Scenario 36	104
4.2.41. Scenario 37	105
4.2.42. Scenario 38	107
4.2.43. Scenario 39	108
4.2.44. Scenario 40	110
4.3. Analysis of Test Results	111
CHAPTER V CONCLUSION.....	115
5.1. Conclusion	115
5.2. Development Suggestions	116
BIBLIOGRAPHY.....	117

LIST OF FIGURES

Figure 3.1. Research Methodology Diagram	22
Figure 3.2. Sample Data Used.....	24
Figure 3.3. Program Design Flowchart	25
Figure 4.1. Dataset Folder	30
Figure 4.2 Code Program to load dataset	30
Figure 4.3.a. Code Program for data preprocessing	31
Figure 4.3.b. Code Program to load dataset and process data	32
Figure 4.4. Combining data and labels	32
Figure 4.5. Running various scenarios to be tested	33
Figure 4.14. Application of K-Means clustering.....	40
Figure 4.15. Application of K-Nearest Neighbor (KNN).....	40
Figure 4.16. Prediction Visualization	41
Figure 4.21. Prediction results for scenario 1.....	44
Figure 4.22. <i>Confusion matrix scenario 1</i>	45
Figure 4.23. Prediction result for scenario 2	46
Figure 4.24. <i>Confusion matrix scenario 2</i>	46
Figure 4.25. Prediction result for scenario 3	47
Figure 4.26. <i>Confusion matrix scenario 3</i>	48
Figure 4.27. Prediction result for scenario 4	49
Figure 4.28. <i>Confusion matrix scenario 4</i>	49
Figure 4.29. Prediction result for scenario 5	50
Figure 4.30. <i>Confusion matrix scenario 5</i>	51
Figure 4.31. Prediction result for <i>scenario 6</i>	52
Figure 4.32. <i>Confusion matrix scenario 6</i>	52
Figure 4.33. Prediction result for scenario 7	53
Figure 4.34. <i>Confusion matrix scenario 7</i>	54
Figure 4.35. Prediction result for scenario 8	55
Figure 4.36. <i>Confusion matrix scenario 8</i>	55
Figure 4.37. Prediction result for scenario 9	56
Figure 4.38. <i>Confusion matrix scenario 9</i>	57

Figure 4.39. Prediction result for scenario 10	58
Figure 4.40. <i>Confusion matrix scenario 10</i>	58
Figure 4.45. Prediction result for scenario 11	61
Figure 4.46. <i>Confusion matrix scenario 11</i>	62
Figure 4.47. Prediction result for scenario 12	63
Figure 4.48. <i>Confusion matrix scenario 12</i>	63
Figure 4.49. Prediction result for scenario 13	64
Figure 4.50. <i>Confusion matrix scenario 13</i>	65
Figure 4.51. Prediction result for scenario 14	66
Figure 4.52. <i>Confusion matrix scenario 14</i>	66
Figure 4.53. Prediction result for scenario 15	67
Figure 4.54. <i>Confusion matrix scenario 15</i>	68
Figure 4.55. Prediction result for scenario 16	69
Figure 4.56. <i>Confusion matrix scenario 16</i>	69
Figure 4.57. Prediction result for scenario 17	70
Figure 4.58. <i>Confusion matrix scenario 17</i>	71
Figure 4.59. Prediction result for scenario 18	72
Figure 4.60. <i>Confusion matrix scenario 18</i>	73
Figure 4.61. Prediction result for scenario 19	74
Figure 4.62. <i>Confusion matrix scenario 19</i>	75
Figure 4.63. Prediction result for <i>scenario 20</i>	76
Figure 4.64. <i>Confusion matrix scenario 20</i>	76
Figure 4.69. Prediction result for <i>scenario 21</i>	79
Figure 4.70. <i>Confusion matrix scenario 21</i>	80
Figure 4.71. Prediction result for <i>scenario 22</i>	81
Figure 4.72. <i>Confusion matrix scenario 22</i>	81
Figure 4.73. Prediction result for scenario 23	82
Figure 4.74. <i>Confusion matrix scenario 23</i>	83
Figure 4.75. Prediction result for scenario 24	84
Figure 4.76. <i>Confusion matrix scenario 24</i>	84
Figure 4.77. Prediction result for scenario 25	85
Figure 4.78. <i>Confusion matrix scenario 25</i>	86

Figure 4.79. Prediction result for scenario 26	87
Figure 4.80. <i>Confusion matrix scenario 26</i>	87
Figure 4.81. Prediction result for scenario 27	88
Figure 4.82. <i>Confusion matrix scenario 27</i>	89
Figure 4.83. Prediction result for scenario 28	90
Figure 4.84. <i>Confusion matrix scenario 28</i>	90
Figure 4.85. Prediction result for scenario 29	91
Figure 4.86. <i>Confusion matrix scenario 29</i>	92
Figure 4.87. Prediction result for scenario 30	93
Figure 4.88. <i>Confusion matrix scenario 30</i>	93
Figure 4.93. Prediction result for scenario 31	96
Figure 4.94. <i>Confusion matrix scenario 31</i>	97
Figure 4.95. Prediction result for scenario 32	98
Figure 4.96. <i>Confusion matrix scenario 32</i>	98
Figure 4.97. Prediction result for scenario 33	99
Figure 4.98. <i>Confusion matrix scenario 33</i>	100
Figure 4.99. Prediction result for scenario 34	101
Figure 4.100. <i>Confusion matrix scenario 34</i>	101
Figure 4.101. Prediction result for scenario 35	102
Figure 4.102. <i>Confusion matrix scenario 35</i>	103
Figure 4.103. Prediction result for scenario 36	104
Figure 4.104. <i>Confusion matrix scenario 36</i>	104
Figure 4.105. Prediction result for scenario 37	105
Figure 4.106. <i>Confusion matrix scenario 37</i>	106
Figure 4.107. Prediction result for scenario 38	107
Figure 4.108. <i>Confusion matrix scenario 38</i>	107
Figure 4.109. Prediction result for scenario 39	108
Figure 4.110. <i>Confusion matrix scenario 39</i>	109
Figure 4.111. Prediction result for scenario 40	110
Figure 4.112. <i>Confusion matrix scenario 40</i>	110

LIST OF TABLES

Table 4.1. Prediction Result Table	111
Tabel 4.2. Prediction Result Table	112

LIST OF NOTATIONS