



UNDERGRADUATE THESIS

APPLICATION OF MULTISURF FOR FEATURE WEIGHTING AT THE RETRIEVAL STAGE OF CASE-BASED REASONING IN EARLY DETECTION OF TYPE 2 DIABETES MELLITUS

AIMEE NATYA RAMADHANI
NPM 22081010159

THESIS ADVISORS

Dr. Faisal Muttaqin, S.Kom., M.T
Eva Yulia Puspaningrum, S.Kom., M.Kom

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

APPROVAL SHEET

APPLICATION OF MULTISURF FOR FEATURE WEIGHTING AT THE RETRIEVAL STAGE OF CASE-BASED REASONING IN EARLY DETECTION OF TYPE 2 DIABETES MELLITUS

By :
AIMEE NATYA RAMADHANI
NPM. 22081010159

It has been defended in front of and accepted by the Thesis Examiner Team of the Informatics Study Program, Faculty of Computer Science, National Development University of East Java Veterans on July 10, 2026:

Dr. Faisal Muttagin, S.Kom., M.T
NIP. 19851231 2021211 009



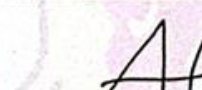
(Thesis Advisors I)

Eva Yulia Puspaningrum, S.Kom., M.Kom
NIP. 19890705 2021212 002



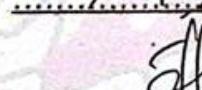
(Thesis Advisors II)

Dr. Eng. Ir. Anggraini Puspita Sari, ST., MT.
NPT. 222198 60 816400



(Head Assessor)

Fetty Tri Anggraeny, S.Kom. M.Kom
NIP. 19820211 2021212 005



(Assessor II)

Acknowledge by,,
Dean of the Faculty of Computer Science



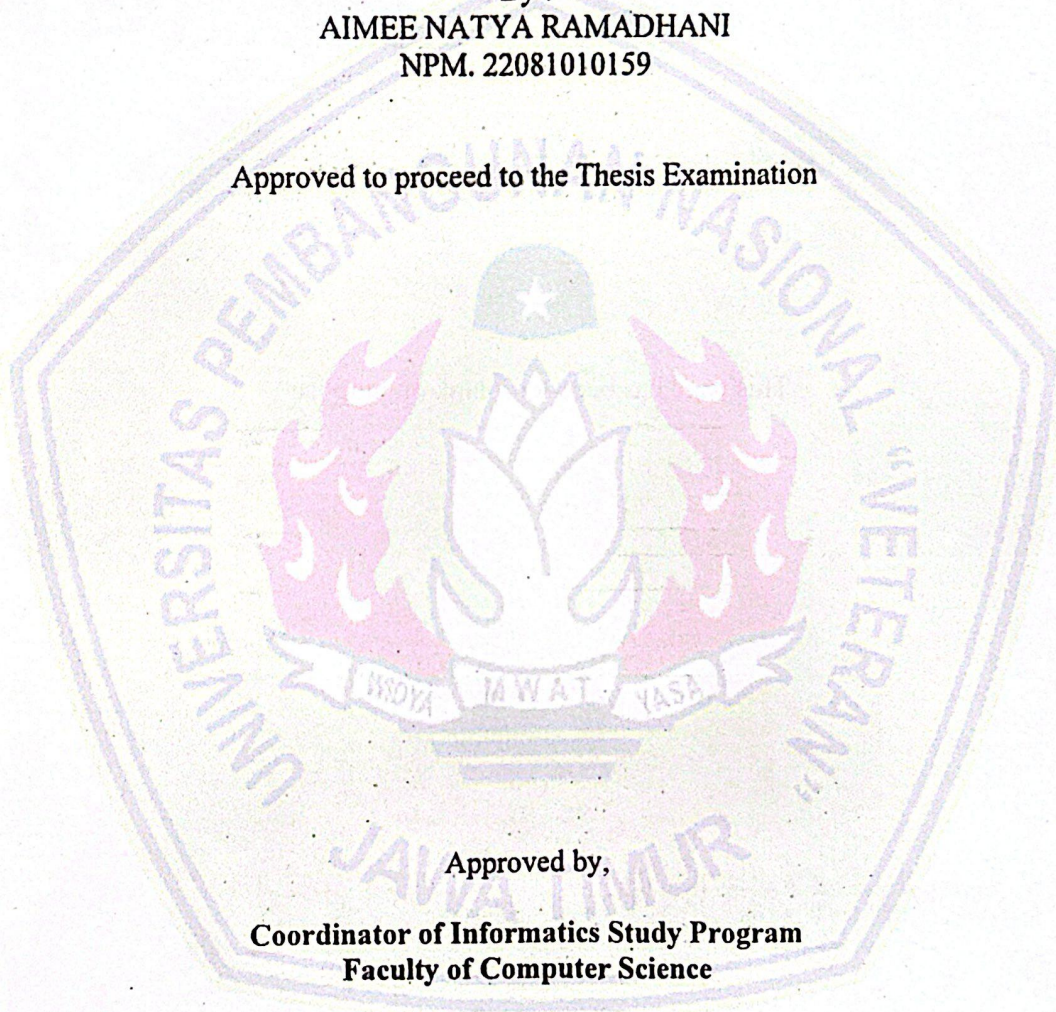
Prof. Dr. Ir. Novirina Hendrasari, MT
NIP. 19681126 199403 2 001

APPROVAL SHEET

APPLICATION OF MULTISURF FOR FEATURE WEIGHTING AT THE RETRIEVAL STAGE OF CASE-BASED REASONING IN EARLY DETECTION OF TYPE 2 DIABETES MELLITUS

By :
AIMEE NATYA RAMADHANI
NPM. 22081010159

Approved to proceed to the Thesis Examination



Approved by,
Coordinator of Informatics Study Program
Faculty of Computer Science

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

STATEMENT OF ORIGINALITY

The undersigned:

Student Name : Aimee Natya Ramadhani
NPM : 22081010159
Degree Program : Bachelor (S1)
Study Programs : Informatics
Faculty : Faculty Of Computer Science
Supervisor : Dr. Faisal Muttaqin, S.Kom., M.T
Eva Yulia Puspaningrum, S.Kom., M.Kom

Hereby declares that this undergraduate thesis contains no part of any other scientific work that has been submitted to obtain an academic degree at any higher education institution. Furthermore, it does not contain any work or opinions previously written or published by others, except for those which are explicitly cited in this thesis and listed completely in references.

And I declare that this scientific document is free from elements of plagiarism. If in the future indications of plagiarism are found in this Thesis, I am willing to accept sanctions in accordance with the applicable laws and regulations.

Thus, I made this statement without any coercion from anyone and to be used as it should.



Surabaya, August 28 2026
Declarant,




AIMEE NATYA RAMADHANI
NPM. 22081010159

ABSTRACT

Student Name / NPM : Aimee Natya Ramadhani / 22081010159
Thesis Title : APPLICATION OF MULTISURF FOR
FEATURE WEIGHTING AT THE RETRIEVAL
STAGE OF CASE-BASED REASONING IN
EARLY DETECTION OF TYPE 2 DIABETES
MELLITUS
Supervisor : 1. Dr. Faisal Muttaqin, S.Kom., M.T
2. Eva Yulia Puspaningrum, S.Kom., M.Kom

Type 2 Diabetes Mellitus is a chronic disease that develops slowly without obvious early symptoms, so early detection is crucial to prevent further complications. This study aims to apply the Case-Based Reasoning (CBR) method with feature weighting using the MultiSURF algorithm and Weighted Euclidean Distance-based similarity measurement at the retrieval stage to support the early detection of Type 2 Diabetes Mellitus, using the Pima Indians Diabetes Dataset of 768 patient data with eight clinical features. The results showed that the Glucose feature obtained the highest weight followed by BMI and Age, with the best configuration at $n_neighbors = 3$ and $K = 9$ which resulted in an Accuracy of 76.03%, Precision 65.87%, Recall 64.54%, and an F1-Score of 64.95%. Compared to the CBR model without weighting (Baseline), the MultiSURF-based Weighted CBR model has succeeded in improving Accuracy, Precision, and F1-Score, so that this method can be a viable alternative in supporting the early detection of Type 2 Diabetes Mellitus based on clinical data.

Keywords : Case-Based Reasoning, MultiSURF, Feature Weighting, Euclidean Distance, Type 2 Diabetes Mellitus

ACKNOWLEDGEMENTS

Praise be to Allah SWT for all His graces, guidance and gifts to the author so that the thesis with the title **"APPLICATION OF MULTISURF FOR FEATURE WEIGHTING AT THE RETRIEVAL STAGE OF CASE-BASED REASONING IN EARLY DETECTION OF TYPE 2 DIABETES MELLITUS"** CAN BE COMPLETED PROPERLY.

In the process of preparing this thesis, the author realized that this thesis could not be composed properly without the help, support, guidance, motivation, and prayers of various parties. Therefore, with all humility, the author expresses his gratitude to all parties who have helped, both directly and indirectly, so that this thesis can be completed properly. In particular, the author would like to thank the:

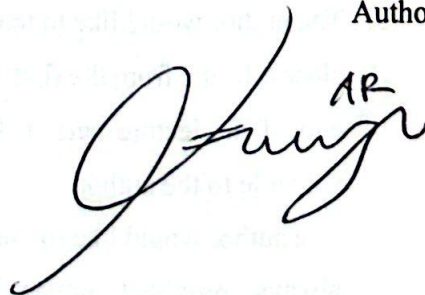
1. The two beloved parents who always provide support, prayers, understanding, patience and an invaluable sense of security, so that the author can complete the study well.
2. Mrs. Prof. Dr. Ir. Novirina Hendrasarie, M.T., as the Dean of the Faculty of Computer Science, East Java Veterans National Development University.
3. Mrs. Dr. Intan Yuniar Purbasari, S.Kom., M.Sc., as the Coordinator of the Informatics Study Program, Faculty of Computer Science, National Development University of East Java Veterans.
4. Mr. Dr. Faisal Muttaqin, S.Kom., M.T. and Mrs. Eva Yulia Puspaningrum, S.Kom., M.Kom., as supervisors who have taken the time and provided guidance, direction, and input during the process of preparing this thesis.
5. The author would like to thank Aileen Anindya Ramadhani, as the writer's closest friend from the first day of re-registration for new students until the end of the lecture period, for all the time and togetherness that is very valuable to the author.
6. The author would like to thank Adelia Deswita and Sativa Freya who have always provided unwavering support from afar. Their presence, encouragement, and kindness have accompanied the author throughout the process, especially during moments when distance made the author feel

alone. Knowing that there were always people who cared, even from afar, made the author feel less alone in facing every challenge along the way.

7. The author would like to thank Ihsanudin Early Adha and Kleyfha Azzahra who have always been there with jokes, advice, and a willingness to listen to the author's endless complaints. Their presence has brought laughter to difficult days and reminded the author that not every problem has to be faced alone. Thank you for being a place to share both the burdens and the little moments that made this journey more bearable.
8. The author would also like to express gratitude to the song *Vienna* by Billy Joel, particularly through the lyrics, "*Slow down, you're doing fine, You can't be everything you wanna be before your time*" which were listened to hundreds of times throughout the completion of this research. These words became a reminder to slow down whenever the fear of being left behind began to arise, and to take a moment to rest and clear the mind without being overwhelmed by anxiety. Throughout this journey, the song quietly reminded the author that there is no need to rush, as everyone has their own pace and time.

The author realizes that in the preparation of this thesis there are many shortcomings. For this reason, constructive criticism and suggestions from all parties are highly expected for the perfection of writing this thesis. Finally, with all the limitations that the author has, hopefully this report can be useful for all parties in general and writers in particular.

Surabaya, 10 July 2026
Author,



AIMEE NATYA RAMADHANI
NPM. 22081010159

TABLE OF CONTENTS

APPROVAL SHEET	i
APPROVAL SHEET	ii
STATEMENT OF ORIGINALITY	Error! Bookmark not defined.
ABSTRACT	iv
ACKNOWLEDGEMENTS.....	v
TABLE OF CONTENTS.....	vi
IMAGE LIST	ix
TABLE LIST	xii
CHAPTER I INTRODUCTION.....	1
1.1 Background	1
1.2. Problem Formulation	3
1.3 Research Objectives	4
1.4 Research Benefits.....	4
1.5 Research Limitations.....	4
CHAPTER II LITERATURE REVIEW	7
2.1 Previous Research	7
2.2 Type 2 Diabetes Mellitus	18
2.3 Case-Based Reasoning.....	19
2.4 Euclidean distance.....	20
2.5 MultiSURF	21
2.6 PIMA Indian Diabetes Patient Dataset	23
2.7 Data Preprocessing.....	24
2.8 Cross validation.....	25
2.9 Min-Max Normalization	25
2.10 SMOTE-ENN.....	26
2.11 Evaluation Metrics	27
2.12 Streamlit	28
CHAPTER III SYSTEM DESIGN AND IMPLEMENTATION.....	29
3.1 Literature Studies	29
3.2 Data Analysis	29

3.3	System Design	32
3.3.1	Data Preprocessing	32
3.3.2	Cross validation 10-Fold.....	34
3.3.3	Feature Weighting with MultiSURF.....	41
3.3.4	Implementation of Case-Based Reasoning	53
3.3.5	Evaluation.....	63
3.4	Determination of the Trial Scenario	65
3.5	System Deployment.....	66
3.6	Website System Design.....	67
CHAPTER IV TESTING AND ANALYSIS.....		71
4.1	Data Preprocessing	71
4.1.1	Dataset Loading and Duplicate Checking.....	71
4.1.2	Missing Value Handling and Feature-Label Separation.....	72
4.2	Median Imputation Results on Missing Value (Practice Data)	73
4.3	Normalizing Data Using MinMaxScaler	75
4.4	Data Balancing Using SMOTE-ENN.....	76
4.5	Feature Weighting with the MultiSURF Algorithm.....	79
4.6	Implementation of Baseline and Weighted CBR	82
4.7	Evaluation of Trial Scenarios	87
4.8	Comparison of Baseline and CBR Weighted Performance	109
4.9	Outcome Analysis.....	112
4.10	System Deployment Results with Streamlit	113
4.10.1	Diabetes Detection and Retrieval Process Page.....	114
4.10.2	Clinical Validation Section (Revise)	115
4.10.3	Case Base Page (Retain).....	116
4.10.4	Analysis of Deployment Results.....	117
CHAPTER V CONCLUSION.....		119
5.1	Conclusion	119
5.2	Suggestions	120
BIBLIOGRAPHY.....		123
APPENDIX.....		129

IMAGE LIST

Figure 2. 1 Condition of Ulcers in Diabetic Feet [34]	19
Figure 2. 2 4R cycles Case-Based Reasoning [7]	20
Figure 3. 1 Research Framework Flow	29
Figure 3. 2 System Design Diagram	32
Figure 3. 3 Flowchart MultiSURF Weighting	41
Figure 3. 4 Flowchart CBR Implementation.....	54
Figure 3. 5 Diagram UseCase	68
Figure 3. 6 Wireframe Detection Page.....	68
Figure 3. 7 Wireframe Estimated Results	69
Figure 3. 8 Wireframe Validation Section	70
Figure 4. 1 Duplicate Data Inspection Results.....	72
Figure 4. 2 Dataset Results After Replacing 0 Values to NaN	73
Figure 4. 3 Median values generated on Fold 1	74
Figure 4. 4 Examples of Data After Median Imputation on Fold 1	74
Figure 4. 5 MinMaxScaler Parameters on Fold1	76
Figure 4. 6 Examples of Data After Normalization	76
Figure 4. 7 Changes in Data Distribution Before and After SMOTE-ENN	78
Figure 4. 8 MultiSURF Average and Standard Weight Deviation at 10 Fold	79
Figure 4. 9 Average Feature Relevance Rate Based on MultiSURF	80
Figure 4. 10 Sample Test Data Profile of Fold 1	84
Figure 4. 11 Summary of Results Voting.....	84
Figure 4. 12 Neighbor Distance Details at K=9.....	85
Figure 4. 13 Confusion Matrix n_Neighbors = 3 on K=1.....	87
Figure 4. 14 Classification Report n_Neighbors = 3 on K=1	88
Figure 4. 15 Confusion Matrix n_Neighbors = 3 on K=3.....	88
Figure 4. 16 Classification Report n_Neighbors = 3 on K=3	89
Figure 4. 17 Confusion Matrix n_Neighbors = 3 on K=5.....	89
Figure 4. 18 Classification Report n_Neighbors = 3 on K=5	90
Figure 4. 19 Confusion Matrix n_Neighbors = 3 on K=7.....	90
Figure 4. 20 Classification Report n_Neighbors = 3 on K=7	91

Figure 4. 21 Confusion Matrix $n_Neighbors = 3$ on $K=9$	91
Figure 4. 22 Classification Report $n_Neighbors = 3$ on $K=9$	92
Figure 4. 23 Confusion Matrix $n_Neighbors = 5$ on $K=1$	92
Figure 4. 24 Classification Report $n_Neighbors = 5$ on $K=1$	93
Figure 4. 25 Confusion Matrix $n_Neighbors = 5$ on $K=3$	93
Figure 4. 26 Classification Report $n_Neighbors = 5$ on $K=3$	94
Figure 4. 27 Confusion Matrix $n_Neighbors = 5$ on $K=5$	94
Figure 4. 28 Classification Report $n_Neighbors = 5$ on $K=5$	95
Figure 4. 29 Confusion Matrix $n_Neighbors = 5$ on $K=7$	95
Figure 4. 30 Classification Report $n_Neighbors = 5$ on $K=7$	96
Figure 4. 31 Confusion Matrix $n_Neighbors = 5$ on $K=9$	96
Figure 4. 32 Classification Report $n_Neighbors = 5$ on $K=9$	97
Figure 4. 33 Confusion Matrix $n_Neighbors = 7$ on $K = 1$	97
Figure 4. 34 Classification Report $n_Neighbors = 7$ on $K = 1$	98
Figure 4. 35 Confusion Matrix $n_Neighbors = 7$ on $K = 3$	98
Figure 4. 36 Classification Report $n_Neighbors = 7$ on $K = 3$	99
Figure 4. 37 Confusion Matrix $n_Neighbors = 7$ on $K = 5$	99
Figure 4. 38 Classification Report $n_Neighbors = 7$ on $K = 5$	100
Figure 4. 39 Confusion Matrix $n_Neighbors = 7$ at $K = 7$	100
Figure 4. 40 Classification Report $n_Neighbors = 7$ at $K = 7$	101
Figure 4. 41 Confusion Matrix $n_Neighbors = 7$ on $K = 9$	101
Figure 4. 42 Classification Report $n_Neighbors = 7$ on $K = 9$	102
Figure 4. 43 Confusion Matrix $n_Neighbors = 9$ on $K = 1$	102
Figure 4. 44 Classification Report $n_Neighbors = 9$ on $K = 1$	103
Figure 4. 45 Confusion Matrix $n_Neighbors = 9$ on $K = 3$	103
Figure 4. 46 Classification Report $n_Neighbors = 9$ on $K = 3$	104
Figure 4. 47 Confusion Matrix $n_Neighbors = 9$ on $K = 5$	104
Figure 4. 48 Classification Report $n_Neighbors = 9$ on $K = 5$	105
Figure 4. 49 Confusion Matrix $n_Neighbors = 9$ on $K = 7$	105
Figure 4. 50 Classification Report $n_Neighbors = 9$ on $K = 7$	106
Figure 4. 51 Confusion Matrix $n_Neighbors = 9$ on $K = 9$	106
Figure 4. 52 Classification Report $n_Neighbors = 9$ on $K = 9$	107

Figure 4. 53 Confusion Matrix CBR Baseline on Best Scenario	111
Figure 4. 54 Pages of Diabetes Risk Detection and Outcomes Retrieval	114
Figure 4. 55 Pages of Diabetes Risk Detection and Outcomes Retrieval	115
Figure 4. 57 Clinical Validation Section (Review).....	115
Figure 4. 58 Case Base Pages (Retain)	116

TABLE LIST

Table 2. 1 Previous Research	11
Table 3. 1 Feature Explanation Pima Indians Diabetes Dataset.....	30
Table 3. 2 Example Table Pima Indians Diabetes Dataset.....	31
Table 3. 3 Examples of Pre-Imputation Data (NaN = Missing).....	33
Table 3. 4 Examples of Results of Separation of Training Data and Test Data	35
Table 3. 5 Results data Median Imputation	36
Table 3. 6 Normalized result data.....	40
Table 3. 7 Final Result of Distance Calculation Euclidean	45
Table 3. 8 Distance Recapitulation and Determination of Data Relevance Status	46
Table 3. 9 Separation Hit and Miss Neighbors.....	48
Table 3. 10 Recapitulations of Δ Difference and Average Glucose & BMI Feature Differences.....	51
Table 3. 11 Examples of MultiSURF Weight Normalization	53
Table 3. 12 Ratings Distance DU1 with DL.....	57
Table 3. 13 Recapitulation of CBR Prediction Results Baseline	58
Table 3. 14 Ratings Distance DU1 with DL.....	62
Table 3. 15 Recapitulation of CBR Prediction Results Weighted.....	63
Table 3. 16 Trial Scenario Table	65
Table 4. 1 CBR Performance Comparison Baseline and CBR Weighted	110