



UNDERGRADUATE THESIS

**PREDICTION OF THE NUMBER OF BLOOD
DONORS USING XGBOOST BASED ON
HISTORICAL AND DEMOGRAPHIC DATA
ANALYSIS (CASE STUDY UDD PMI
BOJONEGORO REGENCY)**

MOCH DANI FERDIAN SAPUTRA
NPM 22081010147

THESIS ADVISORS

Fetty Tri Anggraeny, S.Kom., M.Kom.
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

PREDICTION OF THE NUMBER OF BLOOD DONORS USING XGBOOST BASED ON HISTORICAL AND DEMOGRAPHIC DATA ANALYSIS (CASE STUDY UDD PMI BOJONEGORO REGENCY)

By :
MOCH DANI FERDIAN SAPUTRA
NPM. 22081010147

Has been defended before, and accepted by, the Board of Assessors of the Thesis Examination of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Veteran Jawa Timur, on Juni 12, 2026

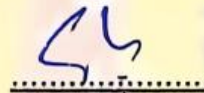
Approved,

Fettv Tri Anggraeny, S.Kom., M.Kom.
NIP. 19820211 202121 2 005



(Advisor I)

Eva Yulia Puspaningrum, S.Kom., M.Kom.
NIP. 19890705 202121 2 002



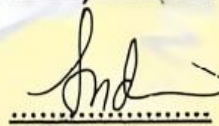
(Advisor II)

Dr. Firza Prima Aditiawan S.Kom., M.T.I,
M.C.F, M.O.S
NIP. 19860523 202121 1 003



(Head Assessor)

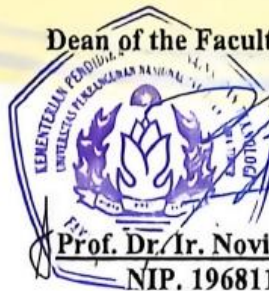
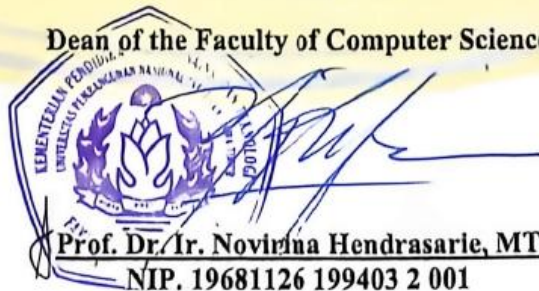
Andreas Nugroho Sihananto, S.Kom., M.Kom.
NIP. 19900412 202406 1 003



(Assessor I)

Acknowledge by,

Dean of the Faculty of Computer Science



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

APPROVAL SHEET

**PREDICTION OF THE NUMBER OF BLOOD DONORS USING
XGBOOST BASED ON HISTORICAL AND DEMOGRAPHIC DATA
ANALYSIS (CASE STUDY UDD PMI BOJONEGORO REGENCY)**

By :
MOCH DANI FERDIAN SAPUTRA
NPM. 22081010147

Approved to proceed to the Thesis Examination



Approved by,

Coordinator of Informatics Study Program
Faculty of Computer Science

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

STATEMENT OF ORIGINALITY

I am the undersigned:

Student Name : Moch Dani Ferdian Saputra
NPM : 22081010147
Degree Program : Bachelor (S1)
Study Program : Informatcs
Faculty : Faculty of Computer Science

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, June 22, 2026
Declarant,



MOCH DANI FERDLAN SAPUTRA
NPM. 22081010147

ABSTRACT

Student Name / NPM : Moch Dani Ferdian Saputra / 22081010147
Thesis Title : Prediction of the Number of Blood Donors Using XGBoost Based on Historical and Demographic Data Analysis (Case Study UDD PMI Bojonegoro Regency)
Advisor : 1. Fetty Tri Anggraeny S.Kom., M.Kom.
2. Eva Yulia Puspaningrum, S.Kom., M.Kom.

Blood supply chain management faces significant challenges in maintaining a balance between stochastic supply and uncertain demand. Supply uncertainty due to various factors, including the impact of the COVID-19 pandemic, highlights the need for an accurate prediction system to optimize blood inventory. This study aims to develop and evaluate a prediction model for the daily and monthly number of blood donors at the Blood Donor Unit (UDD) of the Indonesian Red Cross (PMI) in Bojonegoro Regency using the Extreme Gradient Boosting (XGBoost) algorithm. The data used in the modeling includes 146,447 historical and demographic donor records from January 2019 to December 2024. To achieve better accuracy, this study developed three separate prediction models based on demographic segmentation, namely models based on blood type, gender, and age group. The performance of each model was evaluated at both the daily and monthly aggregate levels using three main metrics: Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and the Coefficient of Determination (R^2). The test results show that the XGBoost algorithm is capable of providing excellent prediction performance across all scenarios. In daily scale predictions, the model successfully captured the fluctuations in the number of donors with R^2 values ranging from 0.755 to 0.974. On a monthly scale, the model's performance increased significantly with R^2 values reaching 0.952 to 0.997. This accuracy improvement proves that the monthly data aggregation process effectively reduces daily fluctuations, resulting in a more stable historical donation pattern. Among the three demographic scenarios tested, the age group-based model, particularly for the 17-24 years age range, proved to provide the most stable and accurate prediction performance. This study concludes that the XGBoost algorithm is highly robust and adaptive in handling the volatility of time-series data on blood donation. It is hoped that this prediction model can serve as a strategic tool for the UDD PMI to proactively plan donor recruitment, optimize inventory management, and mitigate the risks of blood shortages as well as stock wastage.

Keywords : Prediction, Blood Donor, XGBoost, Historical Data, Demographic Data, Machine Learning.

ACKNOWLEDGEMENTS

Praise and thanks be to Allah SWT for all His mercy, guidance, and blessings bestowed upon the author, enabling the successful completion of this thesis titled “Prediction of the Number of Blood Donors Using XGBoost Based on Historical and Demographic Data Analysis (Case Study UDD PMI Bojonegoro Regency)”.

The author would like to express gratitude to all parties involved who have taken the time to provide guidance, advice, and motivation to the author. The author has also received significant assistance from various parties, whether in the form of moral, spiritual, or material support. For this, the author would like to thank:

1. Prof. Dr. Ir. Novirina Hendrasarie, MT, Dean of the Faculty of Computer Science at the “Veteran” National Development University of East Java.
2. Dr. Intan, S.Kom., M.Sc., Head of the Computer Science Program at the Faculty of Computer Science, “Veteran” National Development University of East Java.
3. Mrs. Fetty Tri Anggraeny, S.Kom., M.Kom. and Mrs. Eva Yulia Puspaningrum, S.Kom., M.Kom., as the Academic Advisors who have devoted their time, energy, and thought to guiding, directing, and providing valuable input from the beginning to the completion of this thesis.
4. All the leadership, staff, and members of the Blood Donor Unit (UDD) of the Indonesian Red Cross (PMI) Bojonegoro, who granted research permission, facilitated access, and provided assistance in supplying the data essential for this study.
5. The Examining Professors who provided constructive criticism, suggestions, and evaluations to enhance the quality of this thesis.
6. All the faculty members of the Faculty of Computer Science who have imparted useful knowledge and valuable experience throughout my academic studies.
7. My beloved parents, Father and Mother, for all your sacrifices, love, guidance, and unceasing prayers for the smooth progress of my studies and the completion of this thesis.

8. My dear younger siblings, thank you for your prayers, encouragement, and cheerfulness, which have always been a source of comfort and motivation in completing this final project.
9. My fellow “Eskom Warriors” Alvyn, Mega, and Raihan thank you for your help, time, and companionship through both good times and bad, enabling us to persevere through our academic journey to the very end.
10. My friends from the “Ferrari Youth” Daffa, Desta, Nanta, Abdi, Ivan, Bagas, and Adib thank you for your solidarity, laughter, and the spirit you always shared to help relieve the stress during our college years and while working on this thesis.
11. My friends from “The Boys” Bima and Daffa thank you for your solidarity, support, and the deeply memorable moments of togetherness throughout this academic journey.
12. Friends from the extended family of the BEM Fasilkom UPN “Veteran” East Java, especially to my fellow executives in the AERIAL Cabinet and the Odysseus Cabinet, thank you for the organizational dynamics, collaboration, and valuable experiences that have helped me mature.
13. To someone whose name cannot be mentioned, thank you for all the stories, joys, and sorrows you’ve shared. Although the companionship we thought would last until the very end of our studies actually had to end a few months before graduation, your presence remains an integral part of this journey.

The author acknowledges that this thesis contains many shortcomings. Therefore, constructive criticism and suggestions from all parties are greatly appreciated in order to improve this thesis. Finally, despite the author’s limitations, it is hoped that this report will be of benefit to everyone in general and to the author in particular.

Surabaya, June 12, 2026

Author


Moch Dani Ferdian Saputra

TABLE OF CONTENTS

APPROVAL SHEET	ii
APPROVAL SHEET	iii
STATEMENT OF ORIGINALITY	iv
ABSTRACT	v
ACKNOWLEDGEMENTS.....	vi
TABLE OF CONTENTS.....	viii
LIST OF FIGURES	xi
LIST OF TABLES	xiv
LIST OF ALGORITHMS	xvi
CHAPTER I INTRODUCTION.....	1
1.1 Background	1
1.2 Problem Statement	4
1.3 Research Objective.....	4
1.4 Benefits of Research	5
1.5 Problem Scope	6
CHAPTER II LITERATURE REVIEW	7
2.1 Previous Research	7
2.2 PMI (Indonesian Red Cross Society).....	10
2.3 BSC (Blood Supply Chain).....	11
2.4 Historical Data	11
2.5 Demographic Data	12
2.6 Machine learning.....	12
2.7 Prediction	14
2.8 Regression Model	14
2.9 Ensemble Learning.....	17
2.10 <i>Gradient Boosting</i>	18
2.11 Decision Tree	19
2.12 XGboost	21
2.13 Hyperparameter.....	25
2.14 Mean Absolute Error (MAE)	26
2.15 Root Mean Squared Error (RMSE).....	26

2.16 Coefficient of Determination (R^2).....	27
CHAPTER III RESEARCH METHODOLOGY	28
3.1 Research Phase.....	28
3.2 Problem Identification.....	29
3.3 Data Understanding.....	30
3.4 Data Preprocessing	32
3.4.1 Load Initial Data	32
3.4.2 Data Cleaning and Transformation	33
3.4.3 Exploratory Analysis of Historical and Demographic Data	35
3.4.4 Feature Engineering	40
3.5 Model Development.....	48
3.5.1 Load Data	48
3.5.2 Separation of Input and Target Features	49
3.5.3 Data Distribution	50
3.5.4 Set Hyperparameter.....	50
3.5.5 Algoritm XGBoost.....	51
3.6 Model Evaluation	54
3.6.1 Mean Absolute Error (MAE)	55
3.6.2 Root Mean Squared Error (RMSE).....	56
3.6.3 Koefisien Determinasi (R^2)	56
3.6.4 Visualization of Prediction Results.....	57
3.6.5 Test Scenarios	58
CHAPTER IV RESULTS AND DISCUSSION	61
4.1 Data Preprocessing.....	61
4.1.1 Load Initial Data	61
4.1.2 Data Cleaning and Trnasformation	62
4.1.3 Exploratory Analysis of Historical and Demographic Data	67
4.1.4 Feature Engineering	77
4.2 Model Development.....	82
4.2.1 Load Data	82
4.2.2 Separation of Input and Taget Features	83
4.2.3 Data Distribution.....	87

4.2.4 Set Hyperparameter.....	89
4.2.5 Algoritm XGBoost.....	90
4.3 Model Evaluation.....	110
4.3.1 Daily Scale Model Evaluation.....	110
4.3.2 Monthly Scale Model Evaluation.....	114
4.3.3 Visualization of Daily Scale Prediction Results.....	117
4.3.4 Visualization of Monthly Forecast Results.....	133
4.3.5 Test Scenarios.....	147
CHAPTER V CONCLUSION.....	150
5.1 Conclusions.....	150
5.2 Recommendations.....	151
BIBLIOGRAPHY.....	152
ATTACGMENT 1 Comparison Results of Actual and Predicted.....	156

LIST OF FIGURES

Figure 2. 1 Machine Learning	13
Figure 2. 2 XGBoost Algorithm Scheme	22
Figure 3. 1 Research Methodology.....	28
Figure 3. 2 Example of Visualization of Monthly Donor Trend.....	36
Figure 3. 3 Example of Visualization of Donor Number Trends by Day	36
Figure 3. 4 Example of Monthly Donor Trends by Blood Type.....	37
Figure 3. 5 Example of Daily Donor Trends by Blood Type.....	37
Figure 3. 6 Examples of Monthly Donor Trends by Gender	38
Figure 3. 7 Example of Daily Donor Trends by Gender.....	38
Figure 3. 8 Example of Monthly Donor Trends by Age	39
Figure 3. 9 Example of Daily Donor Trends by Age.....	39
Figure 4. 1 Data Loading Output.....	62
Figure 4. 2 Missing Value Check Output	64
Figure 4. 3 Data Cleaning Output	65
Figure 4. 4 Data Transformation and Merging Output	67
Figure 4. 5 Graph of Monthly Blood Donors.....	69
Figure 4. 6 Graph of Daily Blood Donors	70
Figure 4. 7 Graph of Monthly Donor Analysis by Blood Type.....	72
Figure 4. 8 Graph of Daily Donor Analysis by Blood Type.....	72
Figure 4. 9 Graph of Monthly Donor Analysis by Gender	73
Figure 4. 10 Graph of Daily Donor Analysis by Gender	74
Figure 4. 11 Graph of Monthly Donor Analysis by Age	75
Figure 4. 12 Graph of Daily Donor Analysis by Age	76
Figure 4. 13 Output of Feature Engineering for Blood Type.....	78
Figure 4. 14 Output of Feature Engineering for Gender	80
Figure 4. 15 Output of Feature Engineering for Age	81
Figure 4. 16 Output of Input Feature and Target Splitting for Model 1	85
Figure 4. 17 Output of Input Feature and Target Splitting for Model 2	86
Figure 4. 18 Output of Input Feature and Target Splitting for Model 3	87
Figure 4. 19 Modeling Data Splitting Output	88
Figure 4. 20 First Tree Construction Result of the Model	92

Figure 4. 21 Second Tree Construction Result of the Model.....	93
Figure 4. 22 Loss Curve for Blood Type A	95
Figure 4. 23 Loss Curve for Blood Type B.....	95
Figure 4. 24 Loss Curve for Blood Type AB.....	96
Figure 4. 25 Loss Curve for Blood Type O	96
Figure 4. 26 Loss Curve for Male Gender	97
Figure 4. 27 Loss Curve for Female Gender.....	98
Figure 4. 28 Loss Curve for Age 17-24 Years	98
Figure 4. 29 Loss Curve for Age 25-44 Years	99
Figure 4. 30 Loss Curve for Age 45-70 Years	99
Figure 4. 31 Final Tree for Blood Type A	101
Figure 4. 32 Final Tree for Blood Type B	102
Figure 4. 33 Final Tree for Blood Type AB.....	102
Figure 4. 34 Final Tree for Blood Type O	104
Figure 4. 35 Final Tree for Male Gender	104
Figure 4. 36 Final Tree for Female Gender.....	105
Figure 4. 37 Final Tree for Age 17-24	106
Figure 4. 38 Final Tree for Age 25-44	107
Figure 4. 39 Final Tree for Age 45-70	108
Figure 4. 40 Visualization of Daily Prediction Results by Blood Type.....	118
Figure 4. 41 Visualization of Daily Prediction Results for Blood Type A	119
Figure 4. 42 Visualization of Daily Prediction Results for Blood Type B	120
Figure 4. 43 Visualization of Daily Prediction Results for Blood Type AB	122
Figure 4. 44 Visualization of Daily Prediction Results for Blood Type O	123
Figure 4. 45 Visualization of Daily Prediction Results by Gender.....	124
Figure 4. 46 Visualization of Daily Prediction Results for Male Gender.....	125
Figure 4. 47 Visualization of Daily Prediction Results for Female Gender	126
Figure 4. 48 Visualization of Daily Prediction Results by Age Group.....	128
Figure 4. 49 Visualization of Daily Prediction Results for Age 17-24.....	129
Figure 4. 50 Visualization of Daily Prediction Results for Age 25-44.....	130
Figure 4. 51 Visualization of Daily Prediction Results for Age 45-70.....	132
Figure 4. 52 Visualization of Monthly Prediction Results by Blood Type.....	134

Figure 4. 53 Visualization of Monthly Prediction Results for Blood Type A	135
Figure 4. 54 Visualization of Monthly Prediction Results for Blood Type B	136
Figure 4. 55 Visualization of Monthly Prediction Results for Blood Type AB .	137
Figure 4. 56 Visualization of Monthly Prediction Results for Blood Type O	138
Figure 4. 57 Visualization of Monthly Prediction Results by Gender	139
Figure 4. 58 Visualization of Monthly Prediction Results for Male Gender	140
Figure 4. 59 Visualization of Monthly Prediction Results for Female Gender ..	141
Figure 4. 60 Visualization of Monthly Prediction Results by Age Group	143
Figure 4. 61 Visualization of Monthly Prediction Results for Age 17-24	144
Figure 4. 62 Visualization of Monthly Prediction Results for Age 25-44	145
Figure 4. 63 Visualization of Monthly Prediction Results for Age 45-70	146

LIST OF TABLES

Table 2. 1 Previous Research	7
Table 3. 1 Donor Historical Dataset Example	30
Table 3. 2 Donor Historical Variable Data Type	31
Table 3. 3 Example of Historical Car Unit Dataset	31
Table 3. 4 Historical Variable Data Type Car Unit	32
Table 3. 5 Final Data Example.....	34
Table 3. 6 Variable Data Types in Final Data.....	34
Table 3. 7 Example of Data Based on Blood Type.....	41
Table 3. 8 Features Description in Blood Type Model Data	42
Table 3. 9 Example Data Based on Gender	43
Table 3. 10 Feature Description in Gender Model Data	44
Table 3. 11 Example Data Based on Age	46
Table 3. 12 Feature Description in Age Model Data	47
Table 3. 13 Example of Hyperparameters.....	51
Table 3. 14 Example of XGBoost Implementation Dataset.....	51
Table 3. 15 Example of Gradient and Hessian Calculation	52
Table 3. 16 Example of New Prediction Results	54
Table 3. 17 Example of Evaluation Metric Data.....	55
Table 3. 18 Example of Evaluation Error Calculation.....	55
Table 3. 19 Example of Testing Scenario Table	60
Table 4. 1 Hyperparameters.....	89
Table 4. 2 Evaluation Results of Daily Blood Type Model.....	111
Table 4. 3 Evaluation Results of Daily Gender Model	112
Table 4. 4 Evaluation Results of Daily Age Group Model.....	113
Table 4. 5 Evaluation Results of Monthly Blood Type Model	115
Table 4. 6 Evaluation Results of Monthly Gender Model	115
Table 4. 7 Evaluation Results of Monthly Age Group Model	116
Table 4. 8 Daily Prediction Results of Total Donors by Blood Type	119
Table 4. 9 Daily Prediction Results of Total Donors for Blood Type A	120
Table 4. 10 Daily Prediction Results of Total Donors for Blood Type B.....	121
Table 4. 11 Daily Prediction Results of Total Donors for Blood Type AB.....	122

Table 4. 12 Daily Prediction Results of Total Donors for Blood Type O	123
Table 4. 13 Daily Prediction Results of Total Donors by Gender	124
Table 4. 14 Daily Prediction Results of Total Male Donors.....	126
Table 4. 15 Daily Prediction Results of Total Female Donors	127
Table 4. 16 Daily Prediction Results of Total Donors by Age	128
Table 4. 17 Daily Prediction Results of Total Donors Age 17-24.....	129
Table 4. 18 Daily Prediction Results of Total Donors Age 25-44.....	131
Table 4. 19 Daily Prediction Results of Total Donors Age 45-70.....	132
Table 4. 20 Monthly Prediction Results of Total Donors by Blood Type	134
Table 4. 21 Monthly Prediction Results of Total Donors for Blood Type A.....	135
Table 4. 22 Monthly Prediction Results of Total Donors for Blood Type B.....	136
Table 4. 23 Monthly Prediction Results of Total Donors for Blood Type AB...	137
Table 4. 24 Monthly Prediction Results of Total Donors for Blood Type O.....	139
Table 4. 25 Monthly Prediction Results of Total Donors by Gender	140
Table 4. 26 Monthly Prediction Results of Total Male Donors.....	141
Table 4. 27 Monthly Prediction Results of Total Female Donors	142
Table 4. 28 Monthly Prediction Results of Total Donors by Age	143
Table 4. 29 Monthly Prediction Results of Total Donors Age 17-24	144
Table 4. 30 Monthly Prediction Results of Total Donors Age 25-44	145
Table 4. 31 Monthly Prediction Results of Total Donors Age 45-70	146
Table 4. 32 Testing Scenario.....	147

LIST OF ALGORITHMS

Algorithm 4. 1 Initial Data Loading	61
Algorithm 4. 2 Missing Value Check	63
Algorithm 4. 3 Anomaly Data Cleaning	64
Algorithm 4. 4 Daily Data Aggregation.....	65
Algorithm 4. 5 Merging Demographic and Mobile Unit Data.....	66
Algorithm 4. 6 Overall Monthly Data Analysis.....	68
Algorithm 4. 7 Overall Daily Data Analysis.....	69
Algorithm 4. 8 Analysis of Number of Donors by Blood Type	71
Algorithm 4. 9 Analysis of Number of Donors by Gender.....	73
Algorithm 4. 10 Analysis of Number of Donors by Age Group.....	74
Algorithm 4. 11 Feature Engineering for Blood Type.....	77
Algorithm 4. 12 Feature Engineering for Gender	79
Algorithm 4. 13 Feature Engineering for Age	80
Algorithm 4. 14 Model Development Data Loading	82
Algorithm 4. 15 Feature and Target Splitting	84
Algorithm 4. 16 Data Splitting.....	88
Algorithm 4. 17 Model Training.....	90
Algorithm 4. 18 Daily Model Evaluation	110
Algorithm 4. 19 Monthly Model Evaluation.....	114
Algorithm 4. 20 Visualization of Daily Prediction Results	117
Algorithm 4. 21 Visualization of Monthly Prediction Results	133