UNDERGRADUATE THESIS

COMPARISON OF XGBOOST AND LIGHTGBM METHODS WITH RANDOM OVERSAMPLING FOR OBESITY LEVEL CLASSIFICATION

CHESA SASKIA RAFIKA
NPM 22081010211

THESIS ADVISORS
Eva Yulia Puspaningrum, S.Kom., M.Kom
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

COMPARISON OF XGBOOST AND LIGHTGBM METHODS WITH RANDOM OVERSAMPLING FOR OBESITY LEVEL CLASSIFICATION

By:
CHESA SASKIA RAFIKA
NPM. 22081010211

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 June 18, 2026:

Approved,

Eva Yulia Puspaningrum, S.Kom., M.Kom.
NIP. 19890705 202121 2 002 ..... **(Advisor I)**

Retno Mumpuni, S.Kom., M.Sc.
NIP. 19870716 202521 2 045 ..... **(Advisor II)**

Dr. Eng. Ir. Anggraeni Puspita Sari, ST., MT.
NIP. 222198 60 816400 ..... **(Head Assessor)**

Ardhon Rakhmadi, S.Tr.T., M.Kom.
NIP. 19910805 202406 1 002 ..... **(Assessor I)**

Acknowledge by,

Dean of the Faculty of Computer Science

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

APPROVAL SHEET

COMPARISON OF XGBOOST AND LIGHTGBM METHODS WITH RANDOM OVERSAMPLING FOR OBESITY LEVEL CLASSIFICATION

By:
CHESA SASKIA RAFIKA
NPM. 22081010211

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	:	Chesa Saskia Rafika
NPM	:	22081010211
Degree Program	:	Bachelor (S1)
Study Program	:	Informatics
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, July 6, 2026
Declarant,

CHESA SASKIA RAFIKA
NPM. 22081010211

ABSTRACT

Student Name / NPM	:	Chesa Saskia Rafika / 22081010211
Thesis Title	:	Comparison of XGBoost and LightGBM with Random Oversampling for Obesity Level Classification
Supervisors	:	1. Eva Yulia Puspaningrum, S.Kom., M.Kom
		2. Retno Mumpuni, S.Kom., M.Sc

Obesity is a health issue related to several aspects of individual condition and daily behavior, including physical characteristics, food consumption habits, physical activity, and lifestyle patterns. This study was conducted to evaluate the performance of XGBoost and LightGBM in classifying obesity levels and to analyze the use of Random Oversampling (ROS) for handling unequal class distribution. The research used the Dhaka Obesity Dataset, which initially contained 2,182 respondents with 17 attributes. The data preparation process involved selecting adult respondents by removing data under 18 years old, reconstructing the target labels based on the Asia Pacific Body Mass Index (BMI) classification adopted by the Indonesian Ministry of Health, cleaning the dataset, transforming categorical variables into numerical form, and dividing the data using stratified train test split. After preprocessing and duplicate removal, the final dataset consisted of 2,125 records. The experimental stage was carried out by testing several train test split ratios, applying stepwise manual hyperparameter tuning, and comparing model performance under baseline and ROS scenarios. Model evaluation was performed using accuracy, macro precision, macro recall, macro F1 score, balanced accuracy, confusion matrix, classification report, and train test gap. The results showed that ROS improved the performance of both XGBoost and LightGBM, particularly by producing more balanced classification results across obesity classes. Among all evaluated models, LightGBM with ROS achieved the highest performance, with an accuracy of 0.967059, macro precision of 0.955319, macro recall of 0.942773, macro F1 score of 0.948016, and balanced accuracy of 0.942773. Therefore, LightGBM with ROS was selected as the most suitable model for obesity level classification in this study.

Keywords: Obesity, XGBoost, LightGBM, Random Oversampling, Classification, Class Imbalance.

ACKNOWLEDGEMENTS

All praise and gratitude are directed to Allah SWT for His abundant blessings, guidance, and mercy, which have given the author the strength and opportunity to complete this undergraduate thesis entitled “Comparison of XGBoost and LightGBM with Random Oversampling for Obesity Level Classification.”

This thesis was written as a partial requirement for the completion of the Bachelor’s degree in the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional “Veteran” Jawa Timur. During the preparation and completion of this thesis, the author received valuable assistance, guidance, motivation, and prayers from many parties. The author realizes that this thesis could not have been completed properly without their support. Therefore, the author would like to convey sincere appreciation and gratitude to:

1. The author's beloved parents and family, for their endless love, prayers, encouragement, motivation, and unwavering support throughout the completion of this thesis.
2. Mrs. Eva Yulia Puspaningrum, S.Kom., M.Kom., as the First Supervisor, for her invaluable guidance, constructive suggestions, continuous support, and motivation throughout the research and writing process.
3. Mrs. Retno Mumpuni, S.Kom., M.Sc., as the Second Supervisor, for her guidance, insightful advice, and valuable recommendations that greatly contributed to the completion of this thesis.
4. Mrs. Dr. Eng. Ir. Anggraeni Puspita Sari, S.T., M.T., as the First Examiner, for her constructive criticism, valuable suggestions, and insightful feedback that helped improve the quality of this thesis.
5. Mr. Ardhon Rakhmadi, S.Tr.T., M.Kom., as the Second Examiner, for his valuable guidance, evaluation, and constructive feedback throughout the thesis examination process.
6. All lecturers and academic staff of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur, for their knowledge, guidance, experience, and support throughout the author's academic journey.

7. Dhea Angel Silfany, Vika Rafi Ana, and Revano Maliq Reynanda, for accompanying the author throughout university life, sharing both joyful and challenging moments, and continuously providing encouragement, assistance, and support during the completion of this thesis.
8. Friends and everyone else who cannot be mentioned individually, for their support, encouragement, and assistance throughout the preparation of this thesis.

The author realizes that this thesis still has limitations and may contain shortcomings. Therefore, constructive criticism and suggestions are sincerely welcomed for the improvement of this work in the future. The author hopes that this thesis can provide benefits to readers and serve as a useful reference for future research.

Surabaya, June 19, 2026

Author

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES