**UNDERGRADUATE THESIS**

# APPLICATION OF THE CATBOOST ALGORITHM WITH OPTUNA OPTIMIZATION FOR CLASSIFYING THE ETIOLOGY OF ACUTE FEVER USING SMOTE-NC AS A DATA CLASS BALANCER


**RAIHAN RAMADHAN**
NPM 22081010151


**THESIS ADVISORS**

Yisti Vita Via, S.ST., M.Kom.

Ardhon Rakhmadi, S.Tr.T., 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 THE CATBOOST ALGORITHM WITH OPTUNA OPTIMIZATION FOR CLASSIFYING THE ETIOLOGY OF ACUTE FEVER USING SMOTE-NC AS A DATA CLASS BALANCER

**By:**
RAIHAN RAMADHAN
NPM. 22081010151

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 August 20, 2026:

**Approved,**

| | | |
|---|---|---|
| **Yisti Vita Via, S.ST., M.Kom**<br>NIP. 19860425 202121 2 001 | ............................ | **(Advisor I)** |
| **Ardhon Rakhmadi, S.Tr.T., M.Kom.**<br>NIP. 19910805 202406 1 002 | ............................ | **(Advisor II)** |
| **Fetty Tri Anggraeny, S.Kom, M.Kom**<br>NIP. 19820211 202121 2 005 | ............................ | **(Head Assessor)** |
| **Eva Yulia Puspaningrum, S.Kom., M.Kom**<br>NIP. 19890705 202121 2 002 | ............................ | **(Assessor I)** |

**Acknowledge by,**

**Dean of the Faculty of Computer Science**

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

## APPROVAL SHEET

**APPLICATION OF THE CATBOOST ALGORITHM WITH OPTUNA OPTIMIZATION FOR CLASSIFYING THE ETIOLOGY OF ACUTE FEVER USING SMOTE-NC AS A DATA CLASS BALANCER**

By:
RAIHAN RAMADHAN
NPM. 22081010151

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

I am the undersigned:

| | | |
|---|---|---|
| Student Name | : | Raihan Ramadhan |
| NPM | : | 22081010151 |
| 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, September 9 2026
Declarant,

RAIHAN RAMADHAN
NPM. 22081010151

# ABSTRACT

Student Name / NPM: Raihan Ramadhan/22081010151

Thesis Title: Application of the CatBoost Algorithm with Optuna Optimization for Classifying the Etiology of Acute Fever Using SMOTE-NC as a Data Class Balancer

Supervisor:
1. Yisti Vita Via, S.ST., M.Kom.
2. Ardhon Rakhmadi, S.Tr.T., M.Kom.

This research addresses class imbalance and symptomatic overlap in the multiclass classification of acute fever etiology by integrating CatBoost and SMOTE-NC with hyperparameter optimization via Optuna, implemented as a mobile-based decision support system. The secondary dataset, derived from the AFIRE Study (2013–2016), yielded 801 clean samples across four target classes: Dengue, *Rickettsia spp.*, *Other bacterial*, and *Other viral*. Using Stratified 5-Fold Cross Validation, the optimal CatBoost+SMOTE-NC+Optuna model achieved an average accuracy of 74.16%, Macro Recall of 64.74%, and Macro F1-Score of 64.08%, with the strongest performance on Dengue (F1=0.87) and more moderate results for the minority *Rickettsia spp.* class (F1=0.51). SMOTE-NC significantly improved Macro Recall without a statistically significant loss in Accuracy. Statistical testing confirmed CatBoost significantly outperformed XGBoost across all metrics ($p<0.05$), while showing comparable performance to Random Forest ($p>0.05$); CatBoost was ultimately selected for its architectural efficiency and inference time of just 0.024 ms. A Brier Score-based calibration analysis further confirmed the reliability of the predicted probabilities. The final model was deployed in a client-server architecture (FastAPI backend, Flutter frontend) and successfully passed black-box testing.

**Keywords: CatBoost, SMOTE-NC, Acute Fever, Multiclass Classification, Mobile Application, FastAPI, Flutter**

# ACKNOWLEDGEMENTS

All praise be to Allah SWT for all His blessings, guidance, and grace upon the author, so that this undergraduate thesis titled **"APPLICATION OF THE CATBOOST ALGORITHM WITH OPTUNA OPTIMIZATION FOR CLASSIFYING THE ETIOLOGY OF ACUTE FEVER USING SMOTE-NC AS A DATA CLASS BALANCER"** could be completed properly.

The author would like to express gratitude to Mrs. Yisti Vita Via, S.ST., M.Kom. as Supervisor I and Mr. Ardhon Rakhmadi, S.Tr.T., M.Kom. as Supervisor II, who were willing to take the time to provide guidance, suggestions, and motivation to the author. In addition, during the preparation of the thesis proposal, the author also received a great deal of assistance from various parties. For this, the author would like to thank:

1. Mrs. Prof. Dr. Ir. Novirina Hendrasarie, M.T. as Dean of the Faculty of Computer Science, Universitas Pembangunan Nasional "Veteran" Jawa Timur.
2. Mrs. Dr. Intan Yuniar Purbasari, S.Kom. MSc. as Coordinator of the Informatics Study Program, Faculty of Computer Science, Universitas Pembangunan Nasional Jawa Timur.
3. Mrs. Yisti Vita Via, S.ST. M.Kom., and Mr. Ardhon Rakhmadi, S.Tr.T., M.Kom., as Supervisors who patiently provided direction, guidance, input, and motivation during the process of preparing this thesis.
4. Mrs. Fetty Tri Anggraeny, S.Kom., M.Kom. and Mrs. Eva Yulia Puspaningrum, S.Kom., M.Kom. as the Examining Lecturers who provided constructive criticism, input, and corrections toward the improvement of this thesis.
5. The lecturers of the Informatics Study Program, Universitas Pembangunan Nasional Veteran Jawa Timur, for the guidance provided throughout the author's education.
6. The author's beloved mother, who has always played an important role throughout the author's life journey, always providing support, both material and moral. Always present with her best prayers, becoming a source of strength that never dims, so that the author could persevere to this day.

7. The author's beloved father, who taught the author to be a responsible person, to work hard, and not to give up easily, and whose guidance has always been the author's compass in life.
8. The author's beloved sibling, who has always provided support and input in successfully completing this period of study.
9. My beloved late aunt, Sucik Samiati, and her family, who offered encouragement, welcomed me, and supported me throughout my university studies.
10. dr. Galih Indra Purlistyarini, as the clinical validator who generously provided her time and clinical expertise to validate the research instrument and offer valuable insights from a medical practitioner's perspective.
11. My friends and classmates in the Informatics Study Program, who have struggled together throughout this academic journey, for their companionship, mutual support, and shared perseverance in facing every challenge along the way.
12. My friends at RUMPIGENG, who have also given me support and encouragement and stood by me during both good times and hard times.

The author realizes that there are many shortcomings in the preparation of this thesis. Therefore, constructive criticism and suggestions from all parties are greatly expected for the improvement of this thesis. Finally, with all of the author's limitations, it is hoped that this report can be useful for all parties in general and for the author in particular.

Surabaya, September 3 2026

Raihan Ramadhan

# TABLE OF CONTENTS

# LIST OF FIGURES

# LIST OF TABLES

# LIST OF CODES