IPB University
Bogor Indonesia
@Hak cipta milik IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
IPB University
Perpustakaan IPB University

ADDRESSING CLASS IMBALANCE IN UNDER-5 MORTALITY IN ASEAN: A STUDY OF SMOTEBOOST, RUSBOOST, SMOTE, AND BRF

TAHIRA FULAZZAKY

STATISTICS AND DATA SCIENCE
SCHOOL OF DATA SCIENCE, MATHEMATICS, AND INFORMATICS
IPB UNIVERSITY
BOGOR
2025

@Hak cipta milik IPB University

IPB University

IPB University
Bogor Indonesia

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
   a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
   b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

Perpustakaan IPB University

IPB University
Bogor Indonesia

@Hak cipta milik IPB University

STATEMENT OF THESIS AND SOURCES OF INFORMATION AND COPYRIGHT TRANSFER

As the author, I declare that this thesis entitled "Addressing Class Imbalance in Under-5 Mortality in ASEAN: A Study of SMOTEBoost, RUSBoost, SMOTE, and BRF" is my original work with the guidance of the supervisors and has not been submitted in any form to any university. The source derived or cited from published and unpublished works from other writers has been mentioned in the text and listed in the references at the end of this thesis.

The author assigns the copyright of this thesis to IPB University.

Bogor, January 2025

Tahira Fulazzaky
G1501221024

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

IPB University

Perpustakaan IPB University

IPB University
Bogor Indonesia

RINGKASAN

TAHIRA FULAZZAKY. Menangani Ketidakseimbangan Kelas pada Kematian Balita di ASEAN: Studi SMOTEBoost, RUSBoost, SMOTE, dan BRF pada Kasus Kematian Balita di Asia Tenggara. Dibimbing oleh ASEP SAEFUDDIN dan AGUS MOHAMAD SOLEH.

Penelitian ini bertujuan untuk mengevaluasi efektivitas beberapa teknik pembelajaran mesin dalam menangani masalah ketidakseimbangan kelas dan memprediksi kematian balita di Asia Tenggara. Tujuan utama penelitian ini adalah untuk mengidentifikasi model terbaik di antara SMOTEBoost, RUSBoost, SMOTE-Random Forest (SMOTE-RF), dan Balanced Random Forest (BRF) dalam menangani data dengan ketidakseimbangan kelas. Selain itu, penelitian ini bertujuan untuk menerapkan algoritma yang paling efektif pada dataset dari Asia Tenggara, dengan fokus pada Myanmar, Filipina, dan Kamboja, untuk mengidentifikasi faktor risiko utama yang berkontribusi terhadap kematian balita di wilayah ini.

Penelitian ini menggunakan pendekatan dua fase. Pada fase pertama, 13 dataset dengan tingkat ketidakseimbangan kelas yang bervariasi dianalisis untuk menentukan metode terbaik dalam menangani data yang tidak seimbang. Teknik pembelajaran mesin yang dievaluasi meliputi SMOTEBoost, RUSBoost, SMOTE-Random Forest (SMOTE-RF), dan Balanced Random Forest (BRF). Metode ini dibandingkan dengan model dasar seperti Random Forest (RF) dan AdaBoost. Proses pra-pemrosesan dataset melibatkan pembersihan data, *encoding* label, dan pembagian dataset menjadi subset pelatihan (80%) dan pengujian (20%). Penyesuaian hyperparameter dilakukan menggunakan pencarian acak dengan validasi silang (CV) selama 10 iterasi untuk mengoptimalkan kinerja model. Model dievaluasi berdasarkan akurasi seimbang, *recall*, dan efisiensi komputasi.

Pada fase kedua, algoritma yang berkinerja terbaik berdasarkan fase pertama diterapkan pada data empiris yang diperoleh dari Program Survei Demografis dan Kesehatan (DHS). Data ini fokus pada kematian balita di Asia Tenggara, khususnya di negara-negara seperti Myanmar, Filipina, dan Kamboja, di mana tingkat kematian melebihi target SDG 3.2, yaitu 25 kematian per 1.000 kelahiran hidup. Analisis pentingnya fitur dilakukan untuk mengidentifikasi faktor risiko signifikan yang terkait dengan kematian balita. Faktor-faktor ini mencakup karakteristik ibu, status sosial-ekonomi, dan kondisi lingkungan.

Hasil penelitian ini menunjukkan bahwa Balanced Random Forest (BRF) mengungguli metode lainnya, mencapai akurasi seimbang, recall, dan efisiensi komputasi tertinggi. BRF terbukti sangat efektif untuk dataset dengan kelas yang sangat tidak seimbang dan sedang tidak seimbang, dan kinerjanya tetap stabil di berbagai tingkat ketidakseimbangan kelas. Hal ini menunjukkan bahwa BRF adalah metode yang andal dan kuat dalam menangani dataset yang tidak seimbang, terutama dalam aplikasi terkait kesehatan di mana ketidakseimbangan data merupakan masalah umum. Selain itu, efisiensi komputasi BRF menjadikannya cocok untuk aplikasi berskala besar, terutama di lingkungan dengan sumber daya terbatas.

Analisis pentingnya fitur mengungkapkan beberapa faktor krusial yang berkontribusi terhadap kematian balita di Asia Tenggara. Jumlah total anak yang

@Hak cipta milik IPB University
IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
Perpustakaan IPB University

IPB University
Bogor Indonesia

pernah lahir muncul sebagai prediktor yang paling signifikan, menunjukkan bahwa ukuran keluarga yang lebih besar terkait dengan peningkatan risiko kematian. Usia ibu saat melahirkan dan urutan kelahiran (BORD) juga memainkan peran penting, menyoroti pentingnya kesehatan ibu dan urutan kelahiran. Faktor sosial-ekonomi, seperti indeks kekayaan dan tingkat pendidikan suami/partner, ditemukan memiliki pengaruh signifikan terhadap kematian balita. Selain itu, faktor lingkungan seperti akses terhadap air minum yang bersih sangat penting dalam menentukan tingkat kelangsungan hidup anak, yang menunjukkan betapa pentingnya infrastruktur dasar dalam meningkatkan kesehatan anak.

Sebagai kesimpulan, penelitian ini berhasil mengidentifikasi Balanced Random Forest (BRF) sebagai model yang paling efektif untuk memprediksi kematian balita di Asia Tenggara, khususnya dalam menangani dataset yang tidak seimbang. Penelitian ini juga menyoroti beberapa faktor risiko kunci, seperti usia ibu, urutan kelahiran, status sosial-ekonomi, dan kondisi lingkungan, yang memberikan wawasan berharga untuk intervensi kesehatan masyarakat. Dengan menangani faktor-faktor risiko ini melalui kebijakan yang terfokus, seperti meningkatkan kesehatan ibu, meningkatkan akses pendidikan, dan memperbaiki kondisi kehidupan, strategi kesehatan masyarakat dapat lebih baik disesuaikan untuk mengurangi angka kematian balita di wilayah ini. Penggunaan BRF memastikan penilaian yang kuat dan dapat diandalkan terhadap faktor-faktor prediktor ini, menjadikannya alat yang ampuh untuk analisis data terkait kesehatan di lingkungan dengan keterbatasan sumber daya.

Kata kunci: data tidak seimbang, kematian anak, pembelajaran *ensemble*, pembelajaran mesin

@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

IPB University

Perpustakaan IPB University

IPB University
Bogor Indonesia

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

@Hak cipta milik IPB University

SUMMARY

TAHIRA FULAZZAKY. Addressing Class Imbalance in Under-5 Mortality in ASEAN: A Study of SMOTEBoost, RUSBoost, SMOTE, and BRF. Supervised by ASEP SAEFUDDIN, and AGUS MOHAMAD SOLEH.

This research aimed to evaluate the effectiveness of several machine learning techniques in addressing the challenge of class imbalance and predicting under-five child mortality in Southeast Asia. The primary objective was to identify the best model among SMOTEBoost, RUSBoost, SMOTE-Random Forest (SMOTE-RF), and Balanced Random Forest (BRF) in handling imbalanced class data. Additionally, the research sought to apply the most effective algorithm to datasets from Southeast Asia, focusing on Myanmar, the Philippines, and Cambodia, to identify the key risk factors contributing to under-five child mortality in these regions.

The study followed a two-phase approach. In the first phase, 13 datasets with varying levels of class imbalance were analyzed to determine the best method for handling imbalanced data. The machine learning techniques evaluated included SMOTEBoost, RUSBoost, SMOTE-Random Forest (SMOTE-RF), and Balanced Random Forest (BRF). These methods were compared to baseline models like Random Forest (RF) and AdaBoost. The pre-processing of the datasets involved data cleaning, label encoding, and splitting them into training and testing subsets. Hyperparameter tuning was conducted using randomized search cross-validation to optimize the performance of the models. The models were evaluated based on balanced accuracy, recall, and computational efficiency.

The second phase of the research applied the best-performing algorithm, based on the first phase, to empirical data obtained from the Demographic and Health Surveys (DHS) Program. The data focused on under-five child mortality in Southeast Asia, specifically in countries like Myanmar, the Philippines, and Cambodia, where mortality rates exceed the SDG 3.2 target of 25 deaths per 1,000 live births. Feature importance analysis was conducted to identify significant risk factors associated with under-five child mortality. These factors included maternal characteristics, socio-economic status, and environmental conditions.

The findings from this research revealed that Balanced Random Forest (BRF) outperformed other methods, achieving the highest balanced accuracy, recall, and computational efficiency. BRF was particularly effective for datasets with highly and moderately imbalanced classes, and its performance remained stable across various levels of class imbalance. This indicates that BRF is a reliable and robust method for handling imbalanced datasets, especially in health-related applications where data imbalance is a common issue. Moreover, the computational efficiency of BRF makes it suitable for large-scale applications, particularly in resource-limited environments.

Feature importance analysis revealed several critical factors contributing to under-five child mortality in Southeast Asia. The total number of children ever born emerged as the most significant predictor, indicating that larger family sizes were associated with higher mortality risks. Maternal age at first birth and birth order number (BORD) also played crucial roles, highlighting the importance of maternal health and the sequence of births. Socio-economic factors, such as the wealth index

IPB University

Perpustakaan IPB University

and the education level of the husband/partner, were found to significantly influence child mortality. Furthermore, environmental factors like access to clean drinking water were critical in determining child survival rates, underscoring the importance of basic infrastructure in improving child health outcomes.

In conclusion, this research successfully identified Balanced Random Forest (BRF) as the most effective model for predicting under-five child mortality in Southeast Asia, particularly when dealing with imbalanced datasets. The study also highlighted several key risk factors, such as maternal age, birth order, socio-economic status, and environmental conditions, which provide valuable insights for public health interventions. By addressing these risk factors through targeted policies, such as improving maternal health, increasing access to education, and enhancing living conditions, public health strategies can be better aligned to reduce child mortality in the region. The use of BRF ensures a robust and reliable assessment of these predictors, making it a powerful tool for health-related data analysis in resource-constrained settings.

Keywords: child mortality, ensemble learning, imbalanced data, machine learning

IPB University
Bogor Indonesia
@Hak cipta milik IPB University
IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
Perpustakaan IPB University

IPB University
Bogor Indonesia

@Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

IPB University

Perpustakaan IPB University

© Copyright IPB University, 2025
Law protects this copyright.

It is prohibited to cite all or a part of this work without referring to and mentioning the source. Citation is permitted only for educational purposes, scientific writing, reporting, critical, writing, or reviewing. Citation, whatsoever, shall not inflict detrimental interests on the name of IPB University.

It is prohibited to republish or reproduce all or part of this work, in any form, without written permission from IPB University.

IPB University
Bogor Indonesia
@Hak cipta milik IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
Perpustakaan IPB University
IPB University

ADDRESSING CLASS IMBALANCE IN UNDER-5 MORTALITY IN ASEAN: A STUDY OF SMOTEBOOST, RUSBOOST, SMOTE, AND BRF

TAHIRA FULAZZAKY

Thesis
as a partial requirement for the degree of
Master of Science
in Statistics and Data Science

STATISTICS AND DATA SCIENCE
SCHOOL OF DATA SCIENCE, MATHEMATICS, AND INFORMATICS
IPB UNIVERSITY
BOGOR
2025

External Committee Thesis Examiner:
1 Dr. Bagus Sartono, S.Si., M.Si.

IPB University
Bogor Indonesia
@Hak cipta milik IPB University
IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
Perpustakaan IPB University

IPB University
Bogor Indonesia

Thesis Title : Addressing Class Imbalance in Under-5 Mortality in ASEAN: A Study of SMOTEBoost, RUSBoost, SMOTE, and BRF
Name : Tahira Fulazzaky
Student ID : G1501221024

Approved by

Supervisor 1:
Prof. Dr. Ir. Asep Saefuddin, M.Sc

Supervisor 2:
Dr. Agus Mohamad Soleh, S.Si., M.T

Acknowledged by

Head of Master Program in Statistics and Data Science:
Dr. Agus Mohamad Soleh, S.Si., M.T
NIP 19750315 199903 1 004

Dean of School of Data Science, Mathematics, and Informatics:
Prof. Dr. Ir. Agus Buono, M.Si., M.Kom
NIP 19660702 199302 1 001

Examination Date:
13 January 2025

Graduation Date:

@Hak cipta milik IPB University
IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
Perpustakaan IPB University

IPB University
Bogor Indonesia

ACKNOWLEDGMENT

The author expresses gratitude and praise to Allah for His blessings, enabling the completion of the research proposal titled " Addressing Class Imbalance in Under-5 Mortality in ASEAN: A Study of SMOTEBoost, RUSBoost, SMOTE, and BRF". This research proposal is a requirement for completing a Master of Science degree in Statistics and Data Science program. The writing of this proposal would not have been possible without the support, guidance, and assistance of various individuals, and the author extends thanks to:

1. Prof. Dr. Ir. Asep Saefuddin, M.Sc. and Dr. Agus Mohamad Soleh S.Si., M.T as the supervisory committee members who consistently provided guidance, input, support, and direction to the author;
2. Dr. Bagus Sartono, S.Si., M.Si., as the external examiner, who has provided valuable input and constructive suggestions for the completion of this thesis;
3. The author's father, mother, and brother, who continuously prayed for, supported, and motivated the author;
4. All the lecturers of the Statistics Department for imparting their knowledge;
5. Fellow graduate students of the Statistics and Data Science program, class of 2022, for their support and cooperation.

May this research be beneficial to those in need and contribute to the advancement of knowledge.

Bogor, January 2025

Tahira Fulazzaky

© Hak cipta milik IPB University

Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.

IPB University

Perpustakaan IPB University

IPB University
Bogor Indonesia
@Hak cipta milik IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
IPB University
Perpustakaan IPB University

TABLE OF CONTENTS

LIST OF FIGURES xii

LIST OF TABLES xiii

LIST OF APPENDICES xiii

I INTRODUCTION 1
- 1.1 Background 1
- 1.2 Research Objectives 2

II LITERATURE REVIEW 3
- 2.1 Classification 3
- 2.2 Class Imbalance Problem 3
- 2.3 Imbalance Ratio 3
- 2.4 Resampling Techniques 4
- 2.5 Ensemble Learning 5
- 2.6 Synthetic Minority Oversampling Technique (SMOTE) 6
- 2.7 Random Forest Algoritm 7
- 2.8 Balanced Random Forest 7
- 2.9 RUSBoost 9
- 2.10 SMOTEBoost 10
- 2.11 Under-5 Child Mortality in Southeast Asia 11
- 2.12 Tuning Hyperparameter using Random Search CV 12
- 2.13 Model Evaluation 13
- 2.14 Feature Importance 14

III METHOD 15
- 3.1 Data 15
- 3.2 Analysis Procedure 15

IV RESULTS AND DISCUSSION 18
- 4.1 Evaluations of Models for Handling Imbalanced Data 18
- 4.2 Study on Under 5 Child Mortality in Southeast Asia 31

V CONCLUSION AND RECOMMENDATION 36
- 5.1 Conclusion 36
- 5.2 Recommendation 36

REFERENCES 37

APPENDIX 41

BIOGRAPHY 46

IPB University
Bogor Indonesia

LIST OF FIGURES

1	Random Oversampling (ROS) Process	4
2	Random Undersampling (RUS) Process	4
3	Illustration of SMOTE algorithm	7
4	Conceptual framework of under-5 child mortality	12
5	Flowchart Analysis Procedure	17
6	Balanced accuracy for (a) Brain Stroke and (b) Cerebral Stroke on highly imbalanced datasets	19
7	Recall for (a) Brain Stroke and (b) Cerebral Stroke on highly imbalanced datasets	20
8	Computation time for (a) Brain Stroke and (b) Cerebral Stroke on highly imbalanced datasets	21
9	Balanced accuracy for (a) Fertility, (b) Alpha Thalassemia, (c) Thyroid Cancer, and (d) Cervical Cancer on moderately imbalanced datasets	22
10	Recall for (a) Fertility, (b) Alpha Thalassemia, (c) Thyroid Cancer, and (d) Cervical Cancer on moderately imbalanced datasets	23
11	Computation time for (a) Fertility, (b) Alpha Thalassemia, (c) Thyroid Cancer, and (d) Cervical Cancer on moderately imbalanced datasets	24
12	Balanced accuracy for (a) Diabetes, (b) Monkey Pox, and (c) Early-Stage Diabetes on slightly imbalanced datasets	25
13	Recall for (a) Diabetes, (b) Monkey Pox, and (c) Early-Stage Diabetes on slightly imbalanced datasets	26
14	Computation time for (a) Diabetes, (b) Monkey Pox, and (c) Early-Stage Diabetes on slightly imbalanced datasets	27
15	Balanced accuracy for (a) Screen Time, (b) Breast Cancer, (c) Heart Disease, and (d) Divorce Prediction on balanced datasets	28
16	Recall for (a) Screen Time, (b) Breast Cancer, (c) Heart Disease, and (d) Divorce Prediction on balanced datasets	29
17	Computation time for (a) Screen Time, (b) Breast Cancer, (c) Heart Disease, and (d) Divorce Prediction on balanced datasets	30
18	Age at First Birth and Child Survival Status Across Countries	32
19	Birth Order Frequency Distribution Across Countries (Philippines, Cambodia, Myanmar)	33
20	Distribution of Total Children Ever Born Across Countries	33
21	Feature importance for under-five child mortality in ASEAN	35

@Hak cipta milik IPB University
IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
Perpustakaan IPB University

IPB University
Bogor Indonesia
@Hak cipta milik IPB University
IPB University
Hak Cipta Dilindungi Undang-undang
1. Dilarang mengutip sebagian atau seluruh karya tulis ini tanpa mencantumkan dan menyebutkan sumber :
a. Pengutipan hanya untuk kepentingan pendidikan, penelitian, penulisan karya ilmiah, penyusunan laporan, penulisan kritik atau tinjauan suatu masalah
b. Pengutipan tidak merugikan kepentingan yang wajar IPB University.
2. Dilarang mengumumkan dan memperbanyak sebagian atau seluruh karya tulis ini dalam bentuk apapun tanpa izin IPB University.
Perpustakaan IPB University

LIST OF TABLES

1	The differences between two methodologies: ROS and RUS	5
2	Comparison of Bagging and Boosting in Ensemble Learning	5
3	Comparison between RF with SMOTE and BRF	8
4	Comparison of SMOTEBoost and RUSBoost Algorithm	11
5	Hyperparameters used in BRF, SMOTE-RF, SMOTEBoost and RUSBoost	13
6	Confusion matrix	13
7	List of independent variables (risk factors) to be utilized	15
8	Hyperparameter settings	16
9	Class distribution and imbalance ratios for various health datasets	18
10	Summary of Best Models for Different Datasets	31
11	Class distribution and imbalance ratio (IR) for under-five child mortality across countries	34

LIST OF APPENDICES

1	**Appendix 1** List of datasets and characteristics	42
2	**Appendix 2** Average performance comparison of classification methods across datasets	42
3	**Appendix 3** Model Performance Comparison Across Southeast Asia Datasets	43
4	**Appendix 4** Descriptive Analysis of Under-5 Mortality Variables in Southeast Asia	43
5	**Appendix 5** Authorization Letter from DHS Program	45