



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun  
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

# **POSTPARTUM DEPRESSION DETERMINATION THROUGH MACHINE LEARNING ALGORITHMS**

**WINDA AYU FAZRANINGTYAS**



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun  
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

**SULTAN IDRIS EDUCATION UNIVERSITY**

**2025**



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun  
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun  
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

**POSTPARTUM DEPRESSION DETERMINATION  
THROUGH MACHINE LEARNING  
ALGORITHMS**

**WINDA AYU FAZRANINGTYAS**



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun  
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

**THESIS PRESENTED TO QUALIFY FOR A DOCTOR OF PHILOSOPHY**

**FACULTY COMPUTING AND META-TECHNOLOGY  
SULTAN IDRIS EDUCATION UNIVERSITY**

**2025**



05-4506832



pustaka.upsi.edu.my



Perpustakaan Tuanku Bainun  
Kampus Sultan Abdul Jalil Shah



PustakaTBainun



ptbupsi

LPSI/IPS-3/BO 32  
Pind: 00 m/s: 1/1



Please tick (✓)

Project Paper	<input type="checkbox"/>
Masters by Research	<input type="checkbox"/>
Master by Mixed Mode	<input type="checkbox"/>
PhD	<input checked="" type="checkbox"/>

### INSTITUTE OF GRADUATE STUDIES

### DECLARATION OF ORIGINAL WORK

This declaration is made on the 15 (day) day of April (month) 2025 (year).

i. Student's Declaration:

I, WINDA AYU FAZRANINGTYAS (MATRIC NO. P20211002732), Faculty of Computing and Meta-Technology hereby declare dissertation / thesis for Doctor of Philosophy in Artificial Intelligence titled POSTPARTUM DEPRESSION DETERMINATION THROUGH MACHINE LEARNING ALGORITHMS is my original work. I have not plagiarised from any other scholar's work and any sources that contains copyright had been cited properly for the permitted meanings. Any quotations, excerpt, reference or re-publication from or any works that has copyright had been clearly and well cited.

Signature of the student

ii. Supervisor Declaration:

I, Prof. Dr. Bahbibli Binti Rahmatullah hereby certify that the work entitled POSTPARTUM DEPRESSION DETERMINATION THROUGH MACHINE LEARNING ALGORITHMS was prepared by the above the named student, and was submitted to the Institute of Graduate Studies as a partial fulfilment for the conferment of Doctor of Philosophy in Artificial Intelligence, and the aforementioned work, to the best of my knowledge, is said student's work.

15/04/2025

Date

Professor Dr. Bahbibli Rahmatullah  
Signature of the Supervisor  
Pensyarah

Fakulti Komputeran dan Meta-Teknologi  
Universiti Pendidikan Sultan Idris  
35900 Tanjong Malim, Perak



**INSTITUT PENGAJIAN SISWAZAH /  
INSTITUTE OF GRADUATE STUDIES**

**BORANG PENGESAHAN PENYERAHAN TESIS/DISERTASI/LAPORAN KERTAS PROJEK  
DECLARATION OF THESIS/DISSERTATION/PROJECT PAPER FORM**

Tajuk / Title: POSTPARTUM DEPRESSION DETERMINATION THROUGH MACHINE  
LEARNING ALGORITHMS

No. Matrik / Matric's No.: P20211002732

Saya / I: WINDA AYU FAZRANINGTYAS

(Nama pelajar / Student's Name)

mengaku membenarkan Tesis/Disertasi/Laporan Kertas Projek (Kedoktoran/Sarjana)\* ini disimpan di Universiti Pendidikan Sultan Idris (Perpustakaan Tuanku Bainun) dengan syarat-syarat kegunaan seperti berikut:-

*acknowledged that Universiti Pendidikan Sultan Idris (Tuanku Bainun Library) reserves the right as follows:-*

1. Tesis/Disertasi/Laporan Kertas Projek ini adalah hak milik UPSI.  
*The thesis is the property of Universiti Pendidikan Sultan Idris*
2. Perpustakaan Tuanku Bainun dibenarkan membuat salinan untuk tujuan rujukan dan penyelidikan.  
*Tuanku Bainun Library has the right to make copies for the purpose of reference and research.*
3. Perpustakaan dibenarkan membuat salinan Tesis/Disertasi ini sebagai bahan pertukaran antara Institusi Pengajian Tinggi.  
*The Library has the right to make copies of the thesis for academic exchange.*
4. Sila tandakan ( ✓ ) bagi pilihan kategori di bawah / Please tick ( ✓ ) for category below:-

**SULIT/CONFIDENTIAL**

Mengandungi maklumat yang berdarjah keselamatan atau kepentingan Malaysia seperti yang termaktub dalam Akta Rahsia Rasmi 1972. / Contains confidential information under the Official Secret Act 1972

**TERHAD/RESTRICTED**

Mengandungi maklumat terhad yang telah ditentukan oleh organisasi/badan di mana penyelidikan ini dijalankan. / Contains restricted information as specified by the organization where research was done.

**TIDAK TERHAD / OPEN ACCESS**

  
(Tandatangan Pelajar/ Signature)

Tarikh: 15/04/2025

  
Professor Dr. Bahbib Rahmatullah  
(Tandatangan Penyelia/ Signature of Supervisor)  
& (Nama & Coprasman dan Stempel Rasmi)  
Universiti Pendidikan Sultan Idris  
35900 Tanjong Malim, Perak

Catatan: Jika Tesis/Disertasi ini **SULIT** @ **TERHAD**, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaan dengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai **SULIT** dan **TERHAD**.

Notes: If the thesis is **CONFIDENTIAL** or **RESTRICTED**, please attach with the letter from the organization with period and reasons for confidentiality or restriction.



## ACKNOWLEDGMENT

This thesis is structured in partial fulfillment of the requirements for the degree Doctor of Philosophy in Artificial Intelligence at Sultan Idris Education University (UPSI). This study would not have been possible without the presence and support of persons who, in one way or another, had rendered their invaluable assistance, guidance, and encouragement that made this study possible.

I am profoundly grateful to my supervisor, Professor Dr. Bahbibinti Rahmatullah, for generously imparting her expertise and unwavering support throughout this journey. Her extensive knowledge and exacting standards served as constant inspiration, propelling me toward achieving excellence.

A special thanks is given to the examiners, Ts. Dr. Nazre bin Abdul Rashid; Prof. Dr. Muhammad Suzuri bin Hitam; and Prof. Madya Dr. Shahreen binti Kasim for generously offering their time, thoughtful feedback, and expertise to enhance the quality of this work.

I extend my gratitude to Foundation of Indah Banjarmasin and Sari Mulia University for their financial support in undertaking this research project.

I would like to express my sincere thankful and a deep appreciation to the Dean, Vice Dean, Coordinator of the Offshore Program, faculty, and staff of the Faculty of Computing and Meta-Technology, UPSI.

To my husband, son, family, and friends, your unwavering love, patience, and belief in me gave me the strength to persevere, even when the path was difficult. I could not have accomplished this milestone without you.

Finally, I acknowledge and appreciate the participants of my study who volunteered their time to openly share their experiences, which made this research possible. Their contributions brought a deeper understanding to the field and great personal fulfillment to me.

The guidance and support of everyone acknowledged here are deeply cherished. I am thankful beyond words. This work is dedicated to you all.

To one and all, Terima Kasih!





## ABSTRACT

Postpartum depression (PPD) represents a significant mental health concern that affects the relationship between mothers and their infants worldwide, including in Indonesia. Specifically, in Banjarmasin Indonesia, it has been reported that about 43% of postpartum mothers are diagnosed with the symptoms which become a consideration scenario for scientific study. It has been noted that health institutions in Banjarmasin, Indonesia, are deficient in an effective detection tool, primarily due to the expensive, time-intensive, and insufficient nature of existing methods. The conventional screening approach, which relies on questionnaire instruments, fails to yield a meaningful dataset of PPD that can be analyzed scientifically and accurately. A purposive sampling method was applied, yielding a sample of 317 respondents from healthcare institutions in Banjarmasin City. A hybrid resampling technique combining Bootstrapping and the Synthetic Minority Oversampling Technique (SMOTE) was implemented to address the majority/minority class imbalance. Three experimental approaches were adopted using (1) all Postpartum Depression Risk Factors (PPDRF) features, (2) features selection using statistical tests, and (3) automatic features selection using Relief and Backward Elimination (BE). Four Machine Learning (ML) algorithms which are Naïve Bayes (NB), Logistic Regression (LR), Random Forest (RF), and Adaptive Boosting (AdaBoost) were employed to compare the performance of the PPD determination model. The findings revealed that the consistently identified key PPDRF features are mode of delivery, birth weight, fear of delivery, disposition during pregnancy, family relationships, and support from family and husband. On the other hand, the model of RF integrated with BE demonstrated the highest performance, with an accuracy of 91.62% and an Area Under Curve (AUC) of 0.966. In contrast, the statistical test for RF yielded the lowest result, with an accuracy of 73.38% and an AUC of 0.752. The research demonstrated that an ML-based model is capable of accurately identifying PPD in mothers. The results of this research could assist healthcare institutions, particularly in Banjarmasin Indonesia, in improving their clinical decision-making processes.





## PENENTUAN KEMURUNGAN SELEPAS BERSALIN MELALUI ALGORITMA PEMBELAJARAN MESIN

### ABSTRAK

Kemurungan selepas bersalin (PPD) merupakan satu kebimbangan kesihatan mental yang signifikan dan memberi kesan kepada hubungan antara ibu dan bayi mereka di seluruh dunia, termasuk di Indonesia. Secara khusus, di Banjarmasin, Indonesia, dilaporkan bahawa kira-kira 43% ibu selepas bersalin didiagnosis dengan gejala PPD, yang menjadi asas pertimbangan untuk kajian saintifik. Walau bagaimanapun, institusi kesihatan di Banjarmasin menghadapi kekurangan alat pengesanan yang berkesan, terutamanya disebabkan oleh kaedah sedia ada yang mahal, memakan masa, dan tidak mencukupi. Pendekatan saringan konvensional yang bergantung kepada instrumen soal selidik gagal menghasilkan set data PPD yang bermakna untuk dianalisis secara saintifik dan tepat. Dalam kajian ini, kaedah persampelan bertujuan digunakan dan menghasilkan sampel sebanyak 317 responden dari institusi kesihatan di Kota Banjarmasin. Teknik pensampelan semula hibrid yang menggabungkan kaedah *Bootstrapping* dan *Synthetic Minority Oversampling Technique* (SMOTE) dilaksanakan bagi menangani ketidakseimbangan kelas majoriti/minoriti. Tiga pendekatan eksperimen telah digunakan, iaitu (1) semua ciri Faktor Risiko Kemurungan Selepas Bersalin (PPDRF), (2) pemilihan ciri menggunakan ujian statistik, dan (3) pemilihan ciri automatik menggunakan kaedah *Relief* dan *Backward Elimination* (BE). Empat algoritma Pembelajaran Mesin (ML) iaitu Naïve Bayes (NB), *Logistic Regression* (LR), *Random Forest* (RF), dan *Adaptive Boosting* (AdaBoost) telah digunakan untuk membandingkan prestasi model penentuan PPD. Dapatan kajian menunjukkan bahawa ciri utama PPDRF yang dikenal pasti secara konsisten ialah kaedah bersalin, berat lahir bayi, ketakutan terhadap kelahiran, keadaan emosi semasa kehamilan, hubungan keluarga, serta sokongan daripada keluarga dan suami. Sementara itu, model RF yang diintegrasikan dengan BE menunjukkan prestasi tertinggi dengan ketepatan 91.62% dan kawasan di bawah lengkung (AUC) sebanyak 0.966. Sebaliknya, ujian statistik bagi RF menghasilkan keputusan terendah dengan ketepatan 73.38% dan AUC sebanyak 0.752. Kajian ini membuktikan bahawa model berasaskan ML mampu mengenal pasti PPD dalam kalangan ibu dengan ketepatan yang tinggi. Hasil kajian ini dapat membantu institusi kesihatan, khususnya di Banjarmasin, Indonesia, dalam meningkatkan proses membuat keputusan klinikal mereka.



**TABLE OF CONTENTS**

	<b>Page</b>
<b>DECLARATION OF ORIGINAL WORK</b>	II
<b>DECLARATION OF THESIS</b>	III
<b>ACKNOWLEDGMENT</b>	IV
<b>ABSTRACT</b>	V
<b>ABSTRAK</b>	VI
<b>TABLE OF CONTENTS</b>	VIII
<b>LIST OF TABLES</b>	XVII
<b>LIST OF FIGURES</b>	XX
<b>LIST OF ABBREVIATIONS</b>	XXIII
<b>LIST OF APPENDICES</b>	XXV
<b>CHAPTER 1 INTRODUCTION</b>	1
1.1. Overview	1
1.2. Research Background	2
1.3. Problem Statement	7
1.3.1. Issues Related to PPD Dataset in Banjarmasin	7
1.3.2. Issues Related to Feature in PPD Determination	8
1.3.3. Issues Related to Challenges of ML in PPD Determination	9



1.4. Research Objectives	10
1.5. Research Questions	11
1.6. Research Scope	12
1.7. Research Significance	13
1.8. Operational Definitions	15
1.9. Research Contribution	20
1.10. Thesis Organization	22
1.11. Summary	22
<b>CHAPTER 2 LITERATURE REVIEW</b>	<b>24</b>
2.1. Introduction	24
2.2. Postpartum Depression (PPD)	25
2.2.1. Overview	25
2.2.2. Definition of PPD	27
2.2.3. Etiology of PPD	28
2.2.4. Signs and Symptoms of PPD	29
2.2.5. Consequences of PPD	30
2.3. Dataset in PPD	31
2.4. Edinburg Postnatal Depression Scale (EPDS)	37
2.5. Postpartum Depression Risk Factors (PPDRF) Checklist	41
2.5.1. Sociodemographic Factors	47
2.5.2. Obstetrical Factors	48
2.5.3. Pediatric Factors	48
2.5.4. Psychological Factors	49
2.5.5. Cultural Factors	50

2.5.6. Psychosocial Factors	50
2.5.7. Genetic Factors	51
2.6. Application of Machine Learning (ML) in the Healthcare	51
2.6.1. Overview of Methods	52
2.6.2. Application in Adult Health	53
2.6.3. Application in Mental Health	53
2.6.4. Application in Maternal Health	54
2.7. Deterministic Model in Machine Learning (ML)	56
2.7.1. Modeling of PPD Determination	57
2.7.2. Comprehensive Review of Machine Learning Related to Postpartum Depression (PPD)	60
2.7.3. Best Machine Learning Algorithms for PPD Determination	66
2.7.3.1. Naïve Bayes (NB)	67
2.7.3.2. Logistic Regression (LR)	69
2.7.3.3. Random Forest (RF)	71
2.7.3.4. Adaptive Boosting (AdaBoost)	76
2.8. Enhancement in PPD Deterministic Model	78
2.8.1. Cross Validation (CV)	79
2.8.2. Resampling Methods	81
2.8.2.1. Bootstrapping	82
2.8.2.2. Synthetic Minority Oversampling Technique (SMOTE)	84
2.8.2.3. Hybrid resampling methods: Integration of Bootstrapping and SMOTE	86
2.8.3. Feature Selection (FS)	87

2.8.3.1. Relief	88
2.8.3.2. Backward Elimination (BE)	90
2.9. Correlation Statistical Test: Chi-square Test	91
2.10. RapidMiner	92
2.11. Gaps in PPD Determination	94
2.12. Summary	97
<b>CHAPTER 3 RESEARCH METHODOLOGY</b>	100
3.1. Introduction	100
3.2. Research Design	102
3.3. Participants	103
3.4. Instruments	104
3.5. Research Procedure	110
3.5.1. Phase 1: Preliminary Study	111
3.5.2. Phase 2: Data Collection	111
3.5.3. Phase 3: Deterministic Model	113
3.5.3.1. Methodology of Experiment 1: Deterministic Model Based on All PPDRF Features	115
3.5.3.2. Methodology of Experiment 2: Deterministic Model Based on Selected Features by Correlation Statistical Test	116
3.5.3.3. Methodology of Experiment 3: Deterministic Model Based on Automatic Feature Selection	118
3.5.4. Phase 4: Evaluation	119
3.6. Proposed Model Procedure	120
3.6.1. Data Preparation	120

3.6.2. Data Preprocessing: Hybrid Resampling Methods	122
3.6.3. Corelation Statistical Test: Chi Square ( $\chi^2$ )	125
3.6.4. Feature Selection (FS)	126
3.6.4.1. Relief	126
3.6.4.2. Backward Elimination (BE)	128
3.6.5. <i>k</i> -fold Cross Validation	129
3.6.6. Machine Learning (ML) Algorithms	131
3.6.6.1. Naïve Bayes (NB)	131
3.6.6.2. Logistic Regression (LR)	133
3.6.6.3. Random Forest (RF)	134
3.6.6.4. Adaptive Boosting (AdaBoost)	136
3.6.7. Model Validation	137
3.6.7.1. Internal Validation	138
3.6.7.2. External Validation	141
3.6.8. Method of Internal and External Validation Comparison	143
3.6.9. Performance Metrics	144
3.6.9.1. Accuracy	145
3.6.9.2. Precision	146
3.6.9.3. F measure	146
3.6.9.4. Area Under Curve (AUC)	147
3.6.9.5. Specificity	147
3.6.9.6. Sensitivity	148
3.7. Summary	149
<b>CHAPTER 4 DATA ANALYSIS: PARTICIPANTS' PROFILE</b>	<b>151</b>

4.1. Introduction	151
4.2. Sociodemographic Factors	154
4.3. Obstetrical Factors	161
4.4. Pediatric Factors	167
4.5. Psychological Factors	169
4.6. Cultural Factors	172
4.7. Psychosocial Factors	175
4.8. Genetic Factors	177
4.9. Status of Postpartum Depression (PPD)	179
4.10. Dataset of the Study: All PPDRF Features	182
4.11. Summary	185
<b>CHAPTER 5 RESULT AND DISCUSSION</b>	<b>188</b>

5.1. Introduction	188
5.2. Data Preparation	189
5.3. Data Preprocessing: Hybrid Resampling Methods	190
5.4. Experiment 1: Deterministic Model Based on All PPDRF Features	192
5.4.1. Result of Experiment 1	192
5.4.1.1. Naïve Bayes (NB)	193
5.4.1.2. Logistic Regression (LR)	194
5.4.1.3. Random Forest (RF)	195
5.4.1.4. Adaptive Boosting (AdaBoost)	197
5.4.2. Performance Comparison of Algorithms in Experiment 1	199
5.5. Experiment 2: Deterministic Model Based on Selected Features by Correlation Statistical Test	201

5.5.1. Correlation Statistical Test	201
5.5.1.1. Sociodemographic Characteristics	204
5.5.1.2. Obstetrical Characteristics	208
5.5.1.3. Pediatric Characteristics	214
5.5.1.4. Psychological Characteristics	216
5.5.1.5. Cultural Characteristics	220
5.5.1.6. Psychosocial Characteristics	222
5.5.1.7. Genetic Characteristics	224
5.5.2. Dataset of the Study: PPDRF Selected Features by Correlation Statistical Test	225
5.5.3. Result of Experiment 2	227
5.5.3.1. Naïve Bayes (NB)	227
5.5.3.2. Logistic Regression (LR)	228
5.5.3.3. Random Forest (RF)	230
5.5.3.4. Adaptive Boosting (AdaBoost)	231
5.5.4. Performance Comparison of Algorithms in Experiment 2	233
5.6. Experiment 3: Deterministic Model Based on Automatic Feature Selection	236
5.6.1. Feature Importance	237
5.6.2. Result of Experiment 3: Relief	242
5.6.2.1. Naïve Bayes (NB)	242
5.6.2.2. Logistic Regression (LR)	243
5.6.2.3. Random Forest (RF)	244
5.6.2.4. Adaptive Boosting (AdaBoost)	246
5.6.3. Result of Experiment 3: Backward Elimination (BE)	247

5.6.3.1. Naïve Bayes (NB)	248
5.6.3.2. Logistic Regression (LR)	249
5.6.3.3. Random Forest (RF)	251
5.6.3.4. Adaptive Boosting (AdaBoost)	253
5.6.4. Performance Comparison of Algorithms in Experiment 3	255
5.7. Model Validation	259
5.7.1. Internal Validation	260
5.7.1.1. Internal Model Validation of Experiment 1	260
5.7.1.2. Internal Model Validation of Experiment 2	262
5.7.1.3. Internal Model Validation of Experiment 3	263
5.7.2. External Validation	265
5.7.2.1. External Validation of Experiment 1	266
5.7.2.2. External Validation of Experiment 2	267
5.7.2.3. External Validation of Experiment 3	269
5.8. Performance Comparison among Model Validation	271
5.9. Performance Comparison of Result between Experiment and Prior Study	276
5.10. Discussion	282
5.10.1. RQ 1: What is the appropriate dataset for determining PPD cases of the postpartum women in Banjarmasin?	283
5.10.2. RQ 2: What is the dataset best features in determining PPD cases of postpartum women in Banjarmasin?	286
5.10.3. RQ 3: How to assess the best machine learning algorithm for determining PPD in Banjarmasin dataset?	289

5.10.4. RQ 4: How to evaluate the effectiveness of the best machine learning algorithm model for determining PPD?	292
5.11. Summary	295
<b>CHAPTER 6 CONCLUSION AND RECOMMENDATION</b>	299
6.1. Introduction	299
6.2. Conclusion	300
6.3. Limitation and Recommendation for Future Works	302
<b>REFERENCES</b>	305



## LIST OF TABLES

Table No.		Page
1.1.	Relationship between Research Objectives and Research Questions	12
1.2.	Operational Definition of the Study	15
1.3.	Correlation between Research Objectives and Novel Contribution	21
2.1.	Number of Sample Size and Data Sources of PPD Cases Based on Previous Studies	33
2.2.	Screening Tools and Data Source on the PPD Studies	37
2.3.	Possible Risk Factors of PPD Based on Previous Studies	41
2.4.	ML Method Related to the Occurrence of PPD	60
3.1.	Mind Mapping of the RO, RQ, Methods and Expected Output	101
3.2.	The Categorization of Dependent and Independent Variables from PPDRF Checklist	107
3.3.	Application Parameters: Bootstrapping and SMOTE	124
3.4.	The performance gap between internal and external validation consistency	143
4.1.	Manifestation of the Questions on PPD Represent in EPDS Questionnaire	181
4.2.	Variables or Attributes of the Research and the Scale of the Original Data	183
5.1.	Dataset Comparison Before and After Bootstrapping and SMOTE Implementation	191
5.2.	Experiment 1: NB Implementation	193





5.3.	Experiment 1: LR Implementation	194
5.4.	Experiment 1: RF Implementation	196
5.5.	Experiment 1: AdaBoost implementation	197
5.6.	Chi-square Test on the Significant Relationship between Status of PPD and Participants' Variable	202
5.7.	Dataset Attributes in Experiment 2	226
5.8.	Experiment 2: NB Implementation	228
5.9.	Experiment 2: LR Implementation	229
5.10.	Experiment 2: RF Implementation	230
5.11.	Experiment 2: AdaBoost Implementation	232
5.12.	Value Signs of Selected Features by the Feature Selection Algorithms	237
5.13.	Experiment 3: NB Implementation with Relief Feature Selection	243
5.14.	Experiment 3: LR Implementation with Relief Feature Selection	244
5.15.	Experiment 3: RF Implementation with Relief Feature Selection	245
5.16.	Experiment 3: AdaBoost Implementation with Relief Feature Selection	246
5.17.	Experiment 3: NB Implementation with Backward Elimination Feature Selection	248
5.18.	Experiment 3: LR Implementation with Backward Elimination Feature Selection	250
5.19.	Experiment 3: RF Implementation with Backward Elimination Feature Selection	251
5.20.	Experiment 3: AdaBoost Implementation with Backward Elimination Feature Selection	254
5.21.	Experiment 1: Result of Internal Validation	261
5.22.	Experiment 2: Result of Internal Validation	262
5.23.	Experiment 3: Result of Internal Validation with Relief Feature Selection	263





5.24.	Experiment 3: Result of Internal Validation with Backward Elimination Feature Selection	265
5.25.	Experiment 1: Result of External Validation	266
5.26.	Experiment 2: Result of External Validation	268
5.27.	Experiment 3: Result of External Validation with Relief Feature Selection	269
5.28.	Experiment 3: Result of External Validation with Backward Elimination Feature Selection	270
5.29.	Performance Accuracy Gap of Internal and External Validation	271
5.30.	Performance F1 Score Gap of Internal and External Validation	273
5.31.	Performance AUC Gap of Internal and External Validation	274
5.32.	Comparative Analysis of Performance Metrics from Previous Studies	279





## LIST OF FIGURES

No. Figures		Page
1.1.	Thesis Organization	22
2.1.	Study Workflow and Analytical Strategy of Deterministic Model by Gopalakrishnan et al. (2022)	58
2.2.	Study Workflow and Analytical Strategy of Deterministic Model by Andersson et al. (2021)	59
3.1.	PPDRF Checklist Features	106
3.2.	Research Flow of PPD Prediction	110
3.3.	Framework of the Study	114
3.4.	Model Architecture of PPD Determination Based on All PPDRF Features	115
3.5.	Model Architecture of PPD Determination Based on Selected Features by Correlation Statistical Test	117
3.6.	Model Architecture of PPD Determination Based on Automatic Feature Selection	118
3.7.	Procedure of Input Data Processes in RapidMiner	121
3.8.	Architecture of oversampling method using Bootstrapping and SMOTE	123
3.9.	Procedure of Relief Feature Weight in RapidMiner	127
3.10.	Procedure of Backward Elimination Implementation in RapidMiner	128
3.11.	Cross Validation Application in RapidMiner	130
3.12.	Procedure of NB Implementation in RapidMiner	132





3.13.	Procedure of LR Implementation in RapidMiner	134
3.14.	Procedure of RF Implementation in RapidMiner	135
3.15.	Procedure of AdaBoost Implementation in RapidMiner	137
3.16.	Internal Validation: Visualizing the Dataset Preparation Workflow	139
3.17.	Internal Validation: Model Architecture of the Process	140
3.18.	External Validation: Model Architecture of the Process	141
4.1.	Number of Participants Based on Research Location	153
4.2.	Participants' Profile Based on Sociodemographic Factors in terms of Maternal Age: Early and Late Adolescence, Early and Late Adulthood, and Early Elderly	155
4.3.	Participants' Profile Based on Sociodemographic Factors in terms of Ethnicity	156
4.4.	Participants' Profile Based on Sociodemographic Factors in terms of Marital Status	157
4.5.	Participants' Profile Based on Sociodemographic Factors in terms of Highest Educational Attainment	158
4.6.	Participants' Profile Based on Sociodemographic Factors in terms of Occupation	159
4.7.	Participants' Profile Based on Sociodemographic Factors in terms of Gross Monthly Income	160
4.8.	Participants' Profile Based on Obstetrical Factors in terms of Gravida, Parity, Mode of Delivery, Complication During Pregnancy, and Desire to be Pregnant	163
4.9.	Participants' Profile Based on Obstetrical Factors in terms of Number of Living Children, History of Abortion, Menstrual Problem, Duration of Delivery, and Gestational Age	166
4.10.	Participants' Profile Based on Pediatric Factors	168
4.11.	Participants' Profile Based on Psychological Factors	171
4.12.	Participants' Profile Based on Cultural Factors	174
4.13.	Participants' Profile Based on Psychosocial Factors	176



4.14.	Participants' Profile Based on Genetic Factors	178
4.15.	Participants' Profile Based on Status of PPD	180
5.1.	Performance Comparison of ML Algorithms in Experiment 1	199
5.2.	Performance Comparison of the ML Algorithms in Experiment 2	234
5.3.	Performance Comparison of ML Algorithms in Experiment 3	257
5.4.	Study Workflow and Analytical Strategy of PPD Deterministic Model	278



## LIST OF ABBREVIATIONS

AdaBoost	Adaptive Boosting
Acc.	Accuracy
AI	Artificial Intelligence
AUC	Area Under Curve
BE	Backward Elimination
BMI	Body Mass Index
CDC	Centers for Disease Control and Prevention
CS	Caesarean Section
CV	Cross Validation
DT	Decision Tree
EHR	Electronic Health Record
EPDS	Edinburg Postnatal Depression Scale
FS	Feature Selection
IRLSM	Iteratively Reweighted Least Squares Method
IUFD	Intrauterine Fetal Death
L-BFGS	Limited-memory Broyden–Fletcher–Goldfarb–Shanno
LBW	Low Birth Weight
LR	Logistic Regression





L1	Lasso Regression
L2	Ridge Regression
ML	Machine Learning
MLP	Multi-layer Perceptron
NB	Naïve Bayes
NSVD	Normal Spontaneous Vaginal Delivery
PMS	Premenstrual syndrome
PPD	Postpartum Depression or Postnatal Depression
PPDRF	Postpartum Depression Risk Factors checklist
Prec.	Precision
PRISMA	Preferred Reporting Items for Systematic Reviews and Meta-Analyses
PTSD	Post-traumatic Syndrome Disorder
RF	Random Forest
RO	Research Objective
ROC	Receiver Operating Characteristic
RQ	Research Question
Sens.	Sensitivity
SMOTE	Synthetic Minority Oversampling Technique
SLR	Systematic Literature Review
Spec.	Specificity
SVM	Support Vector Machine
WHO	World Health Organization
XGBoost	Extreme Gradient Boosting
XRT	Extreme Randomized Tree





## LIST OF APPENDICES

- A Letter of Conducting Research
- B Research Ethic Approval
- C Instruments of the Study
- D Dataset: Based on All Features
- E Dataset: Based on Correlational Statistical Model
- F Registration of the Dataset
- G Example of the Full Findings in Experiment 3 with Backward Elimination
- H Research Documentation





## CHAPTER 1

### INTRODUCTION



This chapter outlines the context of the issue (**Section 1.2**) that sets the stage for the subjects explored in this thesis, namely Machine Learning (ML) and Postpartum Depression (PPD). The backdrop to the challenges to be addressed encompasses the author's vested interest in the issues highlighted. Furthermore, it serves as a foundation for shaping the problem statement (**Section 1.3**), delineating research objectives (RO) (**Section 1.4**), and formulating research inquiries (**Section 1.5**).

The RO serve as the cornerstone of any research endeavor. They guide researchers in sourcing and refining data, substantiating the phenomena posited, and ultimately yielding outcomes conducive to comprehending, addressing, and preempting the issues outlined in the research's significance (**Section 1.9**). Furthermore, to





fulfill these objectives, clarity is indispensable, necessitating a delineation that steers the RO of the constructs pertinent to this study, expounded upon in the operational definition (**Section 1.7**). Nonetheless, it is imperative to acknowledge the limitations of this research (**Section 1.8**).

## 1.2. Research Background

Maternal mortality serves as a critical indicator of both access to and the quality of maternal healthcare, as well as the overall health status of women of reproductive age (Schutte et al., 2010). The vulnerability of maternal mortality is significant due to the physiological and psychological changes women undergo during childbirth (Howard & Khalifeh, 2020; Park, Kamaus, & Zhang, 2015). Throughout their lifecycle, women gradually transition from singlehood to marriage and ultimately to motherhood. However, not all women successfully navigate the psychological adjustments during the postpartum period. Hence, adaptation becomes crucial to mitigate potential impacts on the mother and the newborn (Huang et al., 2020).

PPD casts a significant impact not only on children, families, and the immediate environment but also on the mother herself. This condition often leads to strained mother-infant relationships and increases the likelihood of developmental delays in the baby. Left untreated, PPD can detrimentally affect maternal well-being, leading to disruptions in sleeping patterns, eating habits, and daily activities (Bauman et al., 2020; Zhang et al., 2021; Zulfiker, Kabir, Biswas, Nazneen, & Uddin, 2021). However, with





proper intervention, both the mother and the baby stand to benefit (Bauman et al., 2020).

The psychological state is further affected by the physical changes in body shape throughout pregnancy, continuing until childbirth. Postpartum psychological issues may manifest as stress, baby blues, anxiety disorders, and PPD (Obrochta, Chambers, & Bandoli, 2020; Zivoder et al., 2019). PPD is akin to a thief, stripping away the happiness and affection a mother feels and gives to her baby (Beck, 2002). Moreover, mothers grappling with PPD face obstacles in carrying out their daily tasks (American Psychological Association, 2008).

Mortality and disability rates are linked to the mental health challenges faced by postpartum mothers (Lim, 2022). Mental health issues, particularly depression, affect approximately 10% of pregnant women and 13% of women globally (WHO, 2022). Contrastingly, in the United States, as many as 1 in 8 women experience PPD (Centers for Disease Control and Prevention [CDC], 2022). In developing countries, the prevalence of depression among pregnant women stands at about 15.6%, while approximately 19.8% of mothers experience it (World Health Organization [WHO], 2022). This indicates a higher percentage of women experiencing depression in developing countries compared to developed ones or even globally. Additionally, the incidence of PPD tends to be higher in Asian countries than in Western ones (Haque, Nawavar, & Breene, 2015). However, the incidence of PPD in Indonesia remains largely unknown. Fazraningtyas (2020a) stated that around 43% of postpartum mothers admitted to Ulin General Hospital and Dr. H. M. Ansari Saleh General Hospital in





Banjarmasin exhibited moderate to severe levels of depression according to the Edinburg Postnatal Depression Scale (EPDS) assessment. Maternal mental health remains a neglected aspect of healthcare in many Asian countries (Koirala & Chuemchit, 2020). Given the implications of maternal depression, this statistic should be a cause for concern.

Furthermore, in the South Kalimantan Province, particularly in the city of Banjarmasin, the prevalence of PPD remains undisclosed (Fazraningtyas, 2020b). According to a preliminary study conducted in the puerperal wards of Ulin General Hospital and Dr. H.M. Anshari Saleh General Hospital, there has been no assessment of depression among postpartum mothers to facilitate early detection of PPD. This aligns with the findings of Nurbaeti, Syafii, and Lestari (2021), who discovered that PPD screening is not yet implemented in Indonesian healthcare services. This is due to the fact that perinatal mental health policies have not been a priority in the country. Government health initiatives primarily emphasize reducing maternal and infant mortality, placing greater importance on physical health assessments than on mental health.

One strategy to ensure appropriate treatment for mothers is early detection (Alharahsheh & Abdullah, 2021; Amanat et al., 2022; Betts, Kisely, & Alati, 2020; Cai, Wang, Luo, & Wei, 2019; Natarajan et al., 2017; Park et al., 2021; Zogan, Razzak, Jameel, & Xu, 2021). The American College of Obstetricians and Gynecologists [ACOG] (2018) advocates for health services to screen for depression symptoms and anxiety at least once during the perinatal period, and to conduct comprehensive





assessments of mood, emotions, and well-being throughout postpartum visits. Similarly, the WHO and Human Reproduction Programme [HRP] (2022) recommends in-person screening for PPD as part of routine postpartum care. Earls, Yogman, Mattson, and Rafferty (2019) also endorse routine screening for PPD integrated into well-child visits. The adoption of rapidly evolving technology can further support these efforts.

Effective predictive models are believed to assist healthcare workers in identifying and efficiently managing at-risk patients, including those vulnerable to PPD (Saqib, Khan, & Butt, 2021; Shin, Lee, Adeluwa, & Hur, 2020; Xin & Rashid, 2021). This predictive model can be implemented using ML, which has significantly advanced the field of mental health and PPD (Shin et al., 2020; Xin & Rashid, 2021). ML enables more confident predictions at the individual patient level (Liu, Zhang, Guo, & Jiang, 2021), serving as a potent computational tool for analyzing medical datasets and extracting valuable insights (Amit et al., 2021).

Timely intervention plays a crucial role in mitigating mental disorders (Xiang et al., 2020). Clinical models can be promptly employed post-delivery to identify the high risk of PPD, facilitating personalized follow-up and cost-effectiveness (Andersson, Bathula, Iliadis, Walter, & Skalkidou, 2021). ML can also contribute to evidence-based practices by timely detecting mental health issues and depression (Chahar, Dubey, & Narang, 2021; Saqib et al., 2021). Developing ML high-performing ML models is essential for early PPD detection to support healthcare services.





The ML model comprises sophisticated algorithms known for their optimal performance. Li et al. (2022) demonstrated that Random Forest (RF) better result compare to Support Vector Machine (SVM) in analyzing depression among social media users. The same applies to the prediction models used for predicting the occurrence of PPD. Shin et al. (2020) reported that RF achieved the highest Area Under Curve (AUC) performance, trailed closely by SVM. Additionally, RF, SVM, and Adaptive Boosting (AdaBoost) emerged as the top-performing algorithms (Shin et al., 2020; Zulfiker et al., 2021). Furthermore, Naïve Bayes (NB) is a widely utilized algorithm in the realm of mental health issues (Alharahsheh & Abdullah, 2021; Amanat et al., 2022; Andersson et al., 2021; Joshi & Kanoongo, 2022; Li et al., 2022; Liu & Shi, 2022; Natarajan et al., 2017; Shin et al., 2020; Wang, Pathak, & Zhang, 2019; Zogan, Razzak, Wang, Jameel, & Xu, 2022).



This study focuses explicitly on determining PPD using ML alongside a contemporary situational assessment approach during pregnancy, which is subsequently categorized as a risk factor. Consequently, research on PPD in Indonesia still needs to be improved, particularly concerning the utilization of ML development approaches, resulting in a scarcity of PPD datasets.





### 1.3. Problem Statement

This section outlines the challenges and concerns that need to be tackled, which can be viewed from three perspectives: the issues with the PPD dataset in Banjarmasin, the features used in model determination, and the challenges associated with ML.

#### 1.3.1. Issues Related to PPD Dataset in Banjarmasin

Issues with mental health during the postpartum period can manifest during pregnancy and persist after childbirth (Handini & Puspitasari, 2021). Amellia and Kismoyo (2022) noted a lack of specific data on PPD cases in Indonesia. A search on the website of the Ministry of Health of the Republic of Indonesia using keywords related to PPD yielded only information on mild emotional disorders in the National Health Profile. Health data for Indonesia is accessible through the Ministry of Health website. However, the incidence of PPD in Indonesia, particularly in the South Kalimantan, is not currently included in the Basic Health Research (Riskesdas) data. This gap exists due to traditional and time-consuming screening methods, which limit the availability of PPD datasets.

The dataset plays a crucial role in achieving optimal ML performance. To validate the model's accuracy, further investigations using more prominent, more diverse populations and datasets are necessary (Javed et al., 2021; Li et al., 2022; Nemesure, Heinz, Huang, & Jacobson 2021). This study aligns with the





recommendation of Natarajan et al. (2017) for increased data volume and utilization of more sophisticated algorithms to enhance outcomes and enable the potential utilization of ML for PPD determination.

### 1.3.2. Issues Related to Feature in PPD Determination

Gopalakrishnan, Venkataraman, Gururajan, Zhou, and Zhu (2022) urged that further research should assess factors associated with the incidence of PPD. Several studies have utilized PPD risk factors as features, they predominantly focus on sociodemographic factors such as maternal age, race/ethnicity, education, and marital status (Shin et al., 2020; Wang et al., 2019). A potential direction for future research would be to expand the study by incorporating responses from ethnic groups that reflect the national racial demographics (Gopalakrishnan et al., 2022). Furthermore, research indicates that employing various dimensionality reduction algorithms during data processing enhances model performance (Zulfiker et al., 2021). Dimensionality reduction, typically applied during data preprocessing, aims to reduce the number of features (dimensions) without discarding critical information from the dataset (Hediyati & Suartana, 2021). A comprehensive set of features can be leveraged in building robust models. (Zogan et al., 2022).





### 1.3.3. Issues Related to Challenges of ML in PPD Determination

There is a scarcity of predictive models utilizing integrated data that consider the intricate interplay between known and potentially unknown risk factors to identify women at heightened risk of PPD (Hochman et al., 2021; Richter, Fishbain, Richter-Levin, & Okon-Singer, 2021). Andersson et al. (2021) argued that there is currently no effective method to determine women prone to postpartum depressive symptoms. In the context of ML, determining the occurrence of the PPD poses several challenges in achieving optimal, robust, and reliable performance. These challenges include selecting the right features and choosing suitable approaches for building deterministic models that meet expectations. This section provides a brief overview of these challenges.



Research on developing predictive models for PPD cases remains limited. A

systematic literature review (SLR) identified only 16 articles focused on predictive models for PPD between 2017 and 2022, although there is a yearly upward trend. Furthermore, the performance of these models in previous studies has ranged between 78% and 90%. Liu, Hankey, Cao, and Chokka (2021) suggested that ML models can undergo cross-validation (CV) using traditional to ensure the model's generalizability based on individual-level predictions, indicating a promising approach for depression prediction. This gap highlights the need for further research to create a more effective determination model through methods such as model optimization, dimensionality reduction, feature selection, or CV.





Given the challenges outlined above, the authors plan to conduct a robust study on a deterministic model using an ML approach that integrates risk factors as features. This model aims to serve as a reliable tool for determining PPD cases, enabling healthcare professionals to deliver timely interventions to postpartum mothers.

#### 1.4. Research Objectives

A clear research framework is achieved by establishing research objectives (RO). The primary aim of this study is to obtain the best model for determining PPD instances utilizing ML techniques. This endeavor is anticipated to advance technological applications within the healthcare domain.



In particular, the objectives of this study are

1. To establish a PPD dataset based on the risk factors for Banjarmasin postpartum women.
2. To determine the dataset best features for determining PPD cases in Banjarmasin postpartum women.
3. To assess the best machine learning algorithm for determining PPD cases in Banjarmasin postpartum women.
4. To evaluate the effectiveness of the best machine learning algorithm model for determining PPD cases.





## 1.5. Research Questions

To attain the envisioned objectives, it is imperative to formulate Research Questions (RQ) to provide clarity regarding the research direction (Thuan, Drechsler, & Antunes, 2019). Hence, the study was guided by the pursuit of solutions to the following questions:

1. What is the appropriate dataset for determining PPD cases of the postpartum women in Banjarmasin?
2. What is the dataset best features in determining PPD cases of postpartum women in Banjarmasin?
3. How to assess the best machine learning algorithm for determining PPD in Banjarmasin dataset?
4. How to evaluate the effectiveness of the best machine learning algorithm model for determining PPD?

The RQ constitutes a crucial component of the research phase, the initial step in addressing the raised research issues. This RQ is intricately linked with the research objectives, which represent the study's outcomes. **Table 1.1** illustrates the correlation between the RO and the designed RQ in this study.



**Table 1.1***Relationship between Research Objectives and Research Questions*

No.	Research Objectives	Research Questions
1.	To establish a PPD dataset based on the risk factors for Banjarmasin postpartum women.	What is the appropriate dataset for determining PPD cases of the postpartum women in Banjarmasin?
2.	To determine the dataset best features for determining PPD cases in Banjarmasin postpartum women.	What is the dataset best features in determining PPD cases of postpartum women in Banjarmasin?
3.	To assess the best machine learning algorithm for determining PPD cases in Banjarmasin postpartum women.	How to assess the best machine learning algorithm for determining PPD in Banjarmasin dataset?
4.	To evaluate the effectiveness of the best machine learning algorithm model for determining PPD cases.	How to evaluate the effectiveness of the best machine learning algorithm model for determining PPD?

This study will utilize primary data directly sourced from the primary population, namely postpartum mothers, employing a questionnaire previously validated for reliability and validity in Indonesia. The questionnaire will employ the EPDS. Furthermore, the features to be collected in this study was acquired from the Postpartum Depression Risk Factor (PPDRF). The data for this study was gathered from eight public healthcare facilities in Banjarmasin City, comprising two government hospitals, two community health centers, and four midwifery clinics. Private healthcare providers and the respondents' living environments were excluded from the scope of the study.



This study was excluded the participants with comorbidities, specifically mothers with concurrent diseases or disorders alongside the primary concern, such as mothers dealing with depression and hyperthyroidism simultaneously. Only the current pregnancy status was evaluated. The study concentrates on various risk factors associated with PPD, encompassing sociodemographic, obstetric, pediatric, psychological, cultural, psychosocial, and genetic factors, as the foundation for constructing the determination model.

Moreover, this study employed four algorithms to enhance them as determination models for PPD: NB, LR, RF, and AdaBoost. **Chapter 2, Section 2.8** provides a detailed discussion regarding the enhancement of these algorithms. However, the developed algorithms will undergo tuning for optimal performance, thereby contributing to PPD determination.

### 1.7. Research Significance

The findings of this study are deemed necessary for the following individuals or groups of persons:

**Postpartum women.** This study helps postpartum women to detect early signs of depression so that they may possess the right healthcare services to reduce and resolve this depressive condition.





**Family members.** Through this study, families are expected to be more aware of the importance of postpartum screening and support for maternal care after childbirth. Additionally, the psychological changes experienced by the mother during pregnancy to childbirth require the support of the husband and the family so that the mother feels that pregnancy is not a burden in life but the happiness and grace that occur throughout her life.

**Health practitioners.** This study helps to increase knowledge, especially for nurses, about the importance of early screening in postpartum women to know the incidence of PPD so that the provision of nursing care that is given to postpartum women is more comprehensive, not only physically, but also psychologically.



**Healthcare administrators.** This study offers significant insights into the

screening of postpartum women to detect the incidence of PPD. The findings can inform the inclusion of PPD screening as a critical component in postpartum care or assessment agendas. Moreover, the study underscores the importance of monitoring psychological changes in postpartum women, a crucial aspect of healthcare management.

**Hospital staff.** This study is expected provide an input for the provision of services and counseling to postpartum women. However, the support provided by health workers, both nurses, and midwives, can be meaningful information for mothers, so that mothers can get through their postpartum period comfortably and successfully.



**Future researchers.** This study, as one of the references, holds significant potential for further research related to PPD determination models. It can serve as a stepping stone for developing screening instruments for women with PPD in Indonesia. The datasets that will be built in this study can be a rich resource for describing PPD problems in Banjarmasin and the development of the determination model of PPD using other ML algorithms. The future work can build upon these foundations.

## 1.8. Operational Definitions

**Table 1.2** emphasizes the terms that used in the study are hereby defined for clarity.

### *Operational Definition of the Study*

<b>The Terms</b>	<b>Operational Definition</b>
Adaptive Boosting (AdaBoost)	The algorithm that will be developed as part of the determination of PPD which is part of the ensemble technique by using an exponential loss function to obtain optimal performance.
Birth weight	Baby's heaviness measured immediately after completion of first aid in newborns, to be avowed in grams that are classified into 1) very low birth weight i.e., infant weight below 1500 grams, 2) low birth weight belonging to infant weight between 1500 and 2500 grams, and 3) normal baby weight is more than 2500 grams.
Complication during pregnancy	Participants' health problems that arise during the pregnancy period and/or exist before the participant's pregnancy.

*(Continue)*

Table 1.2 (Continue)

<b>The Terms</b>	<b>Operational Definition</b>
Condition of the baby	Refers to the state of the baby born to the participant at birth, i.e., healthy, there is a congenital abnormality or other illness that aggravates the infant's status and is treated at the Neonatal Intensive Care Unit (NICU).
Cross-validation	The performance evaluation method or the performance of the proposed deterministic model developed. In this study, 3, 5, 7, and 10-fold CV is used.
Cultural	The aspects and/or beliefs in life that affect the participant's point of view about the family, and are passed down from generation to generation; it also includes religion and traditions.
Dataset	Data collection obtained from the survey results through the EPDS questionnaire and the PPDRF checklist.
Desire to be pregnant	Classifications are given to the participants to describe their disposition towards pregnancy consisting of whether the participant's pregnancy is desirable and planned or whether the participant's pregnancy is unplanned or part of negligence.
Disposition during pregnancy	The sense of participant's disagreement and/or expression does not accept the pregnancy caused by various reasons.
Duration of labor	The time when the participant feels the pain due to labor contractions for the first time until the fourth stage of delivery is over. It is classified into 1) < 5 hour; 2) 5 – 10 hours; and 3) > 10 hours.
Ethnicity	It is the tribe and culture which refers to the place of origin of the participants.
Family history of depression	It is a profile of depression derived by parents, either from the participant or the husband, seen from paternal and maternal lineage.
Fear of delivery	This is a feeling of anxiety and worry that participants feel in the face of their current delivery.

(Continue)

Table 1.2 (Continue)

<b>The Terms</b>	<b>Operational Definition</b>
Features selection	The data analysis method is used to select features or risk factors that have the potential to cause PPD.
Genetic	Conditions derived from both parents that can affect the state of PPD experienced by the participant include the history of depression in the family.
Gross monthly income	Participants' socio-economic status measured from fixed income in a month according to Minimum Wage of South Kalimantan, Indonesia.
Gestational age	The length of pregnancy period calculated from the first day of the last menstrual period (LMP) that participants undergo to birth expressed in weeks, categorized as 1) preterm where the gestational age is less than 37 weeks; 2) term stating the gestational age of 37-40 weeks; and 3) post-term is the gestational age above 40 weeks.
Gravidity	The classification which shows the number of participant pregnancies, grouped into 1) primigravida or a pregnant participant for the first time, and 2) multigravida is a participant containing for the second time or more.
History of abortion	The record of miscarriage experienced by the participant since the participant got married until the recent delivery.
History of depression	It is the trace of depression that had been suffered by participants before pregnancy or in previous pregnancies.
Knowledge of parenting	It is the information owned and known by the participant on how to care for and nurture a newborn so that the baby grows and develops according to his/her age.
Logistic Regression	The algorithm used for deterministic PPD occurrence involves setting the solver and handling missing values as parameters.
Mode of delivery	This is the manner of infant birth consisting of Normal Spontaneous Vaginal Delivery (NSVD) or Caesarean Section (CS).

(Continue)

Table 1.2 (Continue)

The Terms	Operational Definition
Machine Learning	The method used in the development of the determination model for determining of PPD is through the use of intelligent algorithms, namely NB, LR, RF, and AdaBoost.
Menstrual problems	Conditions experienced by participants during the menstrual cycle can interfere with the health and comfort of participants in daily activities, such as dysmenorrhea, etc. Conditions can be affected by factors such as endometriosis, etc.
Marital status	This is a classification that describes the participant's relationship with others. In this study, marital status is divided into five categories, namely: 1) single, aimed at unmarried participants; 2) married, to a participant who has legitimately established a legal relationship in law and/or religion; 3) divorce, aimed at a participant who has separated from spouses, whether legalized and/or religious; 4) widow, addressed to the participant whose one spouse died; and 5) living together, addressed to a participant living with partners in the absence of legalized and/or religious.
Naïve Bayes	The algorithm developed as part of the determination of PPD is based on Bayes' Theorem, which predicts future opportunities based on previous experience to produce performance according to research objectives.
Obstetric	The events that participants felt during pregnancy until delivery includes gravidity, parity, mode of delivery, desire to be pregnant, complication during pregnancy, number of living children, history of abortion, menstrual problems, duration of labor, and gestational age.
Parity	The number of participant pregnancies that produced live children, is divided into two categories such as 1) primiparous – when a participant is ever giving birth to a baby for the first time, and, 2) multiparous refers to a participant who has given birth several times.
Pediatric	Classification of risk factors refers to aspects of the newborn, including the gender of the baby, birth weight, and condition of the baby.

(Continue)

Table 1.2 (Continue)

The Terms	Operational Definition
Postpartum depression	It is a condition of depression or mood disorder experienced by participants which usually appears after childbirth and is stimulated by several risk factors present around the participant. This condition was assessed using the EPDS.
Psychological	Classification refers to the pleasant and unpleasant characteristics of participants during pregnancy to childbirth consisting of several assessment criteria: 1) fear of delivery, 2) disposition during pregnancy, 3) knowledge of parenting, 4) weight gain and 5) history of depression.
Psychosocial	Classification that describes the relationship between the feelings experienced by the participant during the period of pregnancy to delivery and its relationship to the close person and/or family members and any reciprocal relationship that includes 1) relationship with husband, 2) relationship with the family, and 3) support from husband and family.
Random Forest	The ML algorithm developed for the determination of PPD is a combination of decision tree methods to produce optimal performance.
Relationship with husband	Emotional bonds that exist between participants and husbands are judged by three categories, which are good, good enough, and not good.
Relationship with the family	Connection of emotions involving participants and family members from both participants and husbands is expected to produce support for participants during pregnancy until the participant is able to care for the baby independently categorized into good, good enough, and not good.
Risk factors	It is an aspect of personal behavior and lifestyle, an environmental exposure, or an informed or inherited characteristic that, on the basis of epidemiologic evidence, is known to be associated with health-related conditions.
Weight gain	It is the increase of heaviness experienced by the participant during pregnancy.

(Continue)

Table 1.2 (Continue)

<b>The Terms</b>	<b>Operational Definition</b>
Support	Motivation and endorsement given by the husband or other family members, either from the participants or husbands who are able to produce positive and understanding for the participant, during pregnancy and care for the baby so as not to appear feeling loneliness.
Tradition in family	Rituals of pregnancy and childbirth passed down from generation to generation and were trusted by participants, husbands, or other family members and applied participants in everyday life.

### 1.9. Research Contribution

The objective of this study is to contribute to women's health, particularly in the realm of mental health after childbirth. The primary goal of this thesis is to determine the occurrence of PPD through the implementation of ML algorithms. The data analysis process in this research encompasses various comparison techniques for multiple ML algorithms, involving optimization of data preprocessing, feature selection, and parameter tuning. This approach ensures that the model can effectively determine PPD occurrence based on risk factors obtained directly from the patients.

**Table 1.3***Correlation between Research Objectives and Novel Contribution*

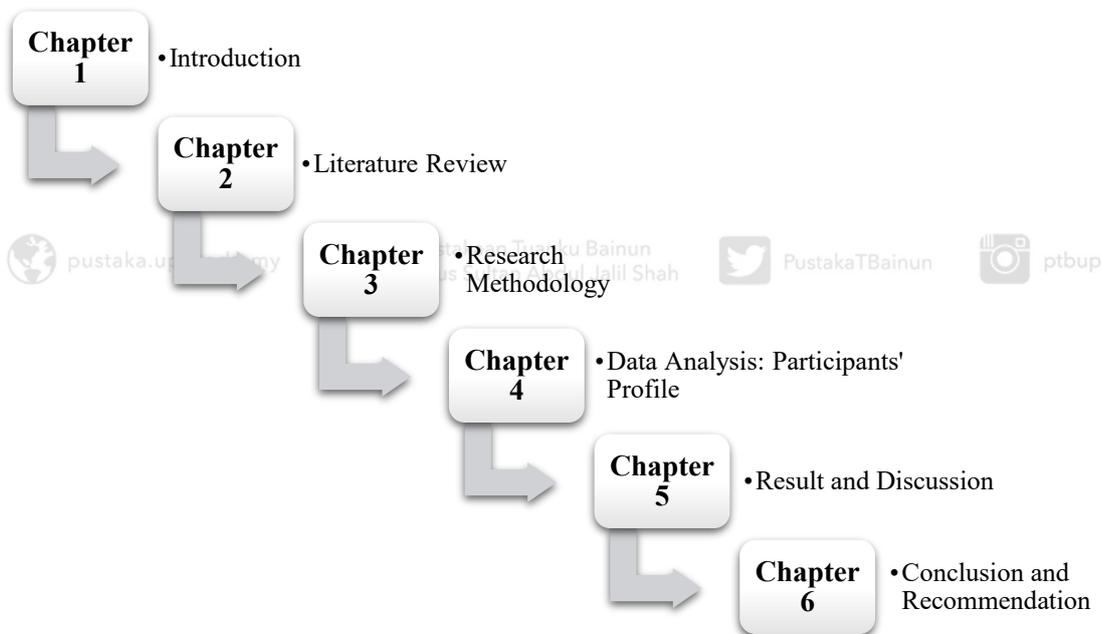
<b>Research Objectives</b>	<b>Novel Contribution</b>
To establish a PPD dataset based on the risk factors for Banjarmasin postpartum women.	<ol style="list-style-type: none"> <li>1. The new dataset was created by utilizing the PPDRF instrument, developed by the literature on risk factors associated with PPD. Furthermore, the assessment of PPD levels employed instruments that underwent testing for both validity and reliability.</li> <li>2. Data for the research were primarily gathered from postpartum mothers who were either hospitalized or attending postpartum care visits at the Banjarmasin City Health Service, South Kalimantan. This information is freshly generated and has yet to be utilized previously.</li> </ol>
To determine the dataset best features for determining PPD cases in Banjarmasin postpartum women.	<ol style="list-style-type: none"> <li>1. This characteristic relies on the inherent risk factors of the participants, aiming to establish a direct connection between the PPD deterministic model and its etiology. The features incorporated in this study distinguish it from prior research, as evidenced by the inclusion of 32 features categorized into seven groups: Sociodemographic Factors (6 features), Obstetrical Factors (10 features), Pediatric Features (3 features), Psychological Factors (5 features), Cultural Factors (2 features), Psychosocial Factors (3 features), Genetic Factors (2 features), with the status of PPD serving as the labels.</li> <li>2. Variables associated with PPD were evaluated using two models: conventional model assessment employing correlation statistics and automated assessment using feature selection. Comparing the effectiveness of these two methods revealed a novel aspect of this research.</li> </ol>
To assess the best machine learning algorithm for determining PPD cases in Banjarmasin postpartum women.	The deterministic model established in this study is executed on four ML algorithms: NB, LR, RF, and AdaBoost. These algorithms serve as tools to determine the occurrence of PPD. This algorithm is enhanced through feature selection with Backward Elimination (BE), Bootstrapping, and SMOTE (as a hybrid model), techniques not employed in prior research, to achieve optimal performance.
To evaluate the effectiveness of the best machine learning algorithm model for determining PPD cases.	

## 1.10. Thesis Organization

This study comprises five interconnected chapters, each detailing the research that will be conducted. The organization of the research serves as a roadmap for structuring the development of the research write-up. **Figure 1.1** depicts the description of each chapter.

**Figure 1.1**

*Thesis Organization*



## 1.11. Summary

This chapter discusses the background of the problem and the justification of the problem in raising the topic of this research. Depression is a mental disorder that generally happens to all levels of age and is a significant contributor to disease



worldwide. One solution to the problem is using algorithms to perform deterministic model with the ML approach. Furthermore, this chapter also talks about the research objectives and the problem formulation which is used as a reference for research development. In addition, this chapter also describes the research scope, motivation, and operational definition, which are closely related to the thesis organization.

This study was conducted against the backdrop of health issues faced by postpartum mothers. It builds upon previous studies by identifying several gaps, thereby emphasizing the necessity for this investigation. These issues intersect with the fields of health and ML. They encompass: 1) maternal depression, a significant public health issue impacting the well-being of both mothers and children, with PPD likely more prevalent than reported; 2) the absence of specific data on PPD cases in Indonesia, as evidenced by a search on the Indonesian Ministry of Health website yielding only information on mild emotional disorders; 3) current detection methods for PPD are costly and lack follow-up, whereas ML offers the potential for more reliable determinations; and 4) optimizing the PPD deterministic model through ML techniques by adjusting parameters and evaluating model effectiveness remains a viable avenue for improvement.

