HYBRID RF-PSO MODEL FOR PREDICTING PREMATURE BIRTH DURING PREGNANCY: A CASE STUDY OF DR. M. ANSARI SALEH HOSPITAL, BANJARMASIN

ANGGRITA SARI

SULTAN IDRIS EDUCATION UNIVERSITY

2025

HYBRID RF-PSO MODEL FOR PREDICTING PREMATURE BIRTH DURING PREGNANCY: A CASE STUDY OF DR. M. ANSARI SALEH HOSPITAL, BANJARMASIN

ANGGRITA SARI

05-4506832 pustaka.upsi.edu.my Perpustakaan Tuanku Bainun Kampus Sultan Abdul Jalil Shah PustakaTBainun ptbupsi

THE THESIS PRESENTED TO QUALIFY FOR A DOCTOR OF PHILOSOPHY

FACULTY OF 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

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

Please tick (√)	
Project Paper	
Masters by Research	
Master by Mixed Mode	
PhD	✓

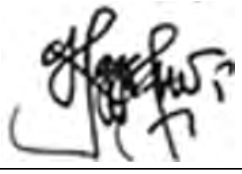
INSTITUTE OF GRADUATE STUDIES

DECLARATION OF ORIGINAL WORK

This declaration is made on the **11th MARCH 2025**

i. Student's Declaration:

I, **ANGGRITA SARI (P20211002733) FACULTY OF COMPUTING AND META-TECHNOLOGY** (PLEASE INDICATE STUDENT'S NAME, MATRIC NO. AND FACULTY) hereby declare that the work entitled **HYBRID RF-PSO MODEL FOR PREDICTING PREMATURE BIRTH DURING PREGNANCY: A CASE STUDY OF DR. M. ANSARI SALEH HOSPITAL, BANJARMASIN** is my original work. I have not copied from any other students' work or from any other sources exceptwhere due reference or acknowledgement is made explicitly in the text, nor has any part been written for me by another person.



Signature of the student

ii. Supervisor's Declaration:

I **ASSOC. PROF MADYA DR. MOHAMMAD MODI BIN LAKULU** (SUPERVISOR'S NAME) hereby certifies that the work entitled **HYBRID RF-PSO MODEL FOR PREDICTING PREMATURE BIRTH DURING PREGNANCY: A CASE STUDY OF DR. M. ANSARI SALEH HOSPITAL, BANJARMASIN** (TITLE) was prepared by the above named student, and was submitted to the Institute of Graduate Studies as a * partial/full fulfillment for the conferment of **DOCTOR OF PHILOSOPHY**, and the aforementioned work, to the best of my knowledge, is the said student'swork.

11 MARCH 2025

Date

Signature of the Supervisor

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

INSTITUT PENGAJIAN SISWAZAH /
INSTITUTE OF GRADUATE STUDIES

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

Tajuk / *Title*: **HYBRID RF-PSO MODEL FOR PREDICTING PREMATURE BIRTH DURING PREGNANCY: A CASE STUDY OF DR. M. ANSARI SALEH HOSPITAL, BANJARMASIN**

No. Matrik /*Matric's No.*: **P20211002733**

Saya / *I* : **ANGGRITA SARI**
(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 kegunaanseperti 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 danpenyelidikan.
   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 pertukaranantara 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 restircted information as specified by the organization where research was done*

☑ **TIDAK TERHAD/*OPEN ACCESS***

(Tandatangan Pelajar/ Signature)

(Tandatangan Penyelia / *Signature of Supervisor*)
& (Nama & Cop Rasmi / *Name & Official Stamp*)

Tarikh: 11 MARCH 2025

Catatan: Jika Tesis/Disertasi ini **SULIT** @ **TERHAD**, sila lampirkan surat daripada pihak berkuasa/organisasi berkenaandengan menyatakan sekali sebab dan tempoh laporan ini perlu dikelaskan sebagai **SULIT** dan **TERHAD**.
Notes: If the thesis is CONFIDENTAL or RESTRICTED, please attach with the letter from the organization with periodand reasons for confidentiality or restriction.

ACKNOWLEDGMENT

After difficulty comes ease. All praises to the Lord of the universe, the owner of all knowledge, for giving the author the ability to complete this thesis, entitled "HYBRID RF-PSO MODEL FOR PREDICTING PREMATURE BIRTH DURING PREGNANCY: A CASE STUDY OF DR. M. ANSARI SALEH HOSPITAL, BANJARMASIN". The author wishes to express sincere gratitude to the following individuals and groups for their invaluable help and support: Prof. Madya Dr. Muhammad Modi bin Lakulu, the first supervisor, for his extensive knowledge, great patience, generous time, motivation, and open collaboration. Dr. Ismail Yusuf Panessai, for his unwavering guidance, directions, and advice in solving every problem encountered during the research. Prof. Madya Dr. Ramlah binti Mailok, for her continual directions and advice in overcoming challenges throughout the research process. All the experts who contributed to the collection of research data and provided information based on their respective fields of expertise. All FCMA staff, for their assistance in problem-solving, contributions of thought, attention, expectations, and prayers throughout the preparation of this thesis. Everyone who has offered help but cannot be mentioned individually—may God grant you double rewards for your support. The author acknowledges that this research has its shortcomings. Therefore, any criticisms, suggestions, and constructive input will be greatly appreciated and will assist the author in further research. Hopefully, this research will contribute to the development of science.

ABSTRACT

Preterm birth (PTB) occurs before 37 weeks of pregnancy. Although many factors influence the course of pregnancy, the actual cause of PTB remains unknown, and current medical efforts focus more on reducing the effects rather than preventing them. Therefore, a technology-based approach is needed to predict essential characteristics and develop a prediction model using machine learning algorithms. This study aims to identify the characteristics of PTB and develop a prediction process model using machine learning. Data was collected from Hospital Dr M. Ansari Saleh in Banjarmasin between 2020 and 2022, involving 915 samples. Factors include maternal factors, demographics, nutritional status, and current pregnancy, which are assessed using a machine- learning algorithm. To determine the best accuracy, this study tested the algorithm's performance with data division ratios of 90/10, 80/20, 70/30, and 60/40. Optimum results were obtained with 80/20 division, then optimized using Particle Swarm Optimization (PSO) and Cross-Validation. The Random Forest algorithm chosen for this study achieved the highest accuracy of 96.45%. Important features identified include the history of abortion, a history of Cesarean surgery (CS), and parity. This research contributes to health informatics by developing an exact PTB prediction model using the Random Forest-PSO hybrid technique. The implication of this study, the model improves health services and can help reduce infant and child deaths due to PTB in Indonesia.

MODEL HIBRID RF- PSO UNTUK MERAMALKAN KELAHIRAN PRAMATANG SEMASA KEHAMILAN KAJIAN KES DI HOSPITAL DR. M ANSARI SALEH DI BANJARMASIN

ABSTRAK

Kelahiran pramatang (PTB) berlaku sebelum 37 minggu kehamilan. Walaupun terdapat banyak faktor yang mempengaruhi perjalanan kehamilan, punca sebenar PTB masih tidak diketahui, dan usaha perubatan semasa lebih menumpukan pada mengurangkan kesan daripada mencegahnya. Oleh itu, pendekatan berasaskan teknologi diperlukan untuk meramalkan ciri-ciri penting dan membangunkan algoritma ramalan yang sesuai. Kajian ini bertujuan untuk mengenal pasti ciri-ciri PTB dan membangunkan model ramalan menggunakan pembelajaran mesin. Data telah dikumpul dari Hospital Dr. M. Ansari Saleh di Banjarmasin antara 2020 dan 2022, melibatkan 915 sampel. Set data yang dipilih termasuk faktor ibu, demografi, status pemakanan dan kehamilan semasa, dinilai menggunakan algoritma pembelajaran mesin. Kajian ini menguji prestasi algoritma dengan nisbah pembahagian data 90/10, 80/20, 70/30, dan 60/40 untuk menentukan ketepatan yang terbaik. Keputusan optimum diperoleh dengan pembahagian 80/20, yang kemudiannya dioptimumkan menggunakan Pengoptimuman Particle Swarm (PSO) dan Pengesahan Silang. Algoritma Random Forest, yang dipilih untuk kajian ini, mencapai ketepatan tertinggi iaitu 96.45%. Ciri penting yang dikenal pasti termasuk sejarah pengguguran, sejarah pembedahan Cesarean (CS), dan pariti. Penemuan ini menyumbang kepada informatika kesihatan dengan membangunkan model ramalan PTB yang sangat tepat menggunakan teknik hibrid Random Forest-PSO. Implikasi kajian ini, model ini dapat menambah baik perkhidmatan kesihatan dan boleh membantu mengurangkan kematian bayi dan kanak-kanak akibat PTB di Indonesia.

CONTENTS

CHAPTER 2 LITERATURE REVIEW

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

AI	Artificial Intelligence
ANC	Antenatal Care
ANN	Artificial Neural Networks
AUC	Area Under the Curve
BMI	Body Mass Index
CART	Classification And Regression Trees
CL	Cervical Length
CNN	Convolutional Neural Networks
CRH	Corticotropin-Releasing Hormone
CV	Cross-Validation
DBN	Deep Belief Networks
DBNs	Deep Trust Networks
DL	Deep Learning
DNN	Deep Neural Networks
DSN	Deep Stacking Network
DT	Decision Tree
FFN	Fetal Fibronectin
FN	False Negative
FP	False Positive
GBest	Global Best

GRU	Gated Recurrent Units
Hb	Hemoglobin
HPA	Hypothalamic-Pituitary-Adrenal
IMR	Infant Mortality Rate
IUGR	Intrauterine Growth Restriction
KIA	Buku Kesehatan Ibu Dan Anak
KNN	K-Nearest Neighbors
KPD	Maternal And Child Health Book
LBW	Premature Rupture of Membranes
LOOCV	Long Short-Term Memory
LSTM	Low Birth Weight
MCH	Multinomialnaïvebayes
ML	Leave-One-Out Cross-Validation
MMP	Maternal Health Care
MNB	Machine Learning
NB	Matrix Metalloproteinase
NLP	Naïve Bayes
P4K	Perencanaan Persalinan Dan Pencegahan Komplikasi
PBest	Premature Rupture of Membranes
PRE	Natural Language Processing
PROM	Premature Birth
PSO	Precision
PTB	Particle Swarm Optimization
PTB	Personal Best

RF	Recurrent Neural Networks
RNNs	Childbirth Planning and Complication Prevention
ROC	Recurrent Neural Networks
SC	Systemic Lupus Erythematosus
SDGs	Sexually Transmitted Infections
SENS	Rank Order Centroid
SLE	Specificity
SOM	Random Forest
SPEC	Gated Recurrent Units (Gru), Self-Organizing Maps
STIs	Sensitivity
TN	Caesarean Section
TP	Sustainable Development Goals
UNISEF	True Positive
USG	True Negative
VLBW	United Nations Children's Fund
WHO	Ultrasonokagrafi

LIST OF APPENDICES

CHAPTER 1

INTRODUCTION

1.1 Introduction

Premature birth (PTB) is a critical public health issue, recognized as the leading cause of neonatal mortality and a significant contributor to long-term health complications in surviving infants (Modell et al., 2022). The urgency of developing effective predictive models to identify pregnant women at high risk of PTB has garnered substantial attention in recent years. Machine learning algorithms have emerged as powerful tools in this domain, enabling healthcare providers to implement timely interventions that could mitigate the risks associated with PTB (Robinson & Norwitz, 2021). By leveraging diverse data sources, these models aim to enhance clinical decision-making, ultimately improving mothers' and babies' health outcomes.

The effectiveness of PTB prediction hinges on the quality and comprehensiveness of the underlying data. Factors such as maternal health history, demographic characteristics, and clinical indicators are crucial in influencing the likelihood of PTB (Li et al., 2021). However, researchers often face significant challenges related to incomplete and inconsistent datasets, which can lead to inaccurate predictions. Consequently, identifying relevant features contributing to PTB risk is essential for refining predictive models. Accurate feature selection is vital for enhancing the predictive power of machine learning algorithms, as it helps focus on the most impactful variables and reduces noise within the data.

This study employs a HYBRID RF-PSO model to address the complexities associated with PTB prediction, which integrates Random Forest (RF) and Particle Swarm Optimization (PSO) techniques. The RF algorithm is known for its robustness and accuracy in classification tasks, mainly when dealing with high-dimensional datasets (Zhao et al., 2023). PSO, inspired by the social behaviour of birds, is a powerful optimization technique that can effectively identify optimal feature subsets and improve the performance of machine learning models (Iqbal et al., 2024). Combining these two methodologies enables enhanced feature selection and model optimization, potentially leading to more accurate predictions of PTB.

The HYBRID RF-PSO model's ability to optimize feature selection is particularly relevant in PTB prediction, where numerous interrelated risk factors exist. By utilizing PSO to search for optimal feature subsets, researchers can identify which factors most significantly contribute to the likelihood of PTB (Sun et al., 2022). This process improves the accuracy of predictions but also aids in understanding the

relationships between various risk factors, thereby facilitating better clinical insights into PTB risk management. Furthermore, the hybrid approach reduces the computational burden often associated with high-dimensional data analysis, enabling faster model training and evaluation.

In addition to improving prediction accuracy, the HYBRID RF-PSO model addresses the need for scalable solutions applicable across diverse populations and clinical settings. This adaptability is essential, as PTB risk factors may vary significantly among different demographic groups (Hershey et al., 2022). The model aims to provide a valuable tool for healthcare providers across different regions and healthcare systems by demonstrating its applicability in various contexts. The insights gained from applying this model can inform targeted interventions and resource allocation, ultimately enhancing prenatal care practices.

Another critical benefit of employing the HYBRID RF-PSO model is its potential to inform preventive measures for at-risk populations. By accurately identifying high-risk individuals, healthcare providers can tailor their approaches to include more intensive monitoring and support for pregnant women. This proactive stance can significantly reduce the incidence of complications associated with PTB, ultimately improving maternal and neonatal health outcomes (Fernandez et al., 2024). The implications of this research extend beyond immediate clinical benefits, potentially leading to reduced healthcare costs associated with managing PTB complications.

This study aims to identify the features contributing to PTB during pregnancy and develop a predictive model utilizing the HYBRID RF-PSO algorithm. The

objective is to analyze the model's accuracy to assist healthcare providers in implementing early preventive and intervention measures, potentially saving lives and improving the health of mothers and babies. By enhancing methods and increasing access to relevant data, this research seeks to advance the accuracy and utility of PTB predictions in clinical practice, ensuring broad applicability and effectiveness across diverse populations and pregnancy characteristics.

1.2 Research Background

Premature birth (PTB) refers to the birth of a baby before 37 weeks of gestation, categorized into subgroups based on gestational age: extremely preterm (less than 28 weeks), very preterm (28 to less than 32 weeks), and moderate to late preterm (32 to 37 weeks). In contrast, term births occur between the 37th and 42nd weeks of pregnancy, and births after 42 weeks are called post-term (Nsugbe et al., 2021). PTB can arise from spontaneous preterm labour or medical interventions, such as planned early labour induction or cesarean sections. In 2020, approximately 13.4 million babies were born prematurely, accounting for over 10% of births worldwide. Furthermore, around 900,000 children died in 2019 due to complications related to prematurity, which can lead to lifelong disabilities, including learning and sensory impairments.

The global health implications of PTB are significant, as it is the leading cause of death in children under five years old. Survival rates, however, vary dramatically by region. In low-income countries, about half of the babies born at or before 32 weeks of gestation die due to inadequate medical care, including insufficient breastfeeding

support and poor management of respiratory difficulties (Modell et al., 2022). In contrast, nearly all babies born in high-income countries survive, benefiting from advanced healthcare services. Middle-income countries often experience challenges where premature babies may face disabilities due to suboptimal utilization of available medical technology. This disparity underscores the need for targeted interventions to improve care for at-risk populations.

Numerous factors contribute to preterm birth, with the majority of cases occurring spontaneously. These factors can include infections, pregnancy complications, multiple gestations, and chronic conditions like diabetes and hypertension. Genetic predispositions and early labour induction or cesarean births also affect PTB occurrences (Visser et al., 2020). Despite the identification of many risk factors, numerous cases remain without a clearly defined cause, highlighting the complexity of PTB as a public health challenge, with around 15 million child deaths attributed to it annually (Surendian et al., 2022; WHO, 2019). Organizations like UNICEF emphasize the necessity for preventive strategies, advocating for primary interventions to mitigate PTB risks during pregnancy (Despotovic et al., 2018).

In high-income countries, reductions in premature death rates can be attributed to policy changes aimed at enhancing healthcare access and quality. These initiatives often involve improving prenatal care, adopting evidence-based practices, and promoting maternal health throughout pregnancy. By prioritizing these measures, significant decreases in PTB incidence and related mortality have been achieved (Pari et al., 2018). However, ongoing efforts are needed to address the remaining challenges, particularly in regions with higher PTB rates.

In Indonesia, the Basic Health Research report for 2022 indicated that 18,281 children under five years old died, with a majority of these deaths occurring during the neonatal period (0-28 days). Notably, 75.5% of these deaths transpired within the first week of life. South Kalimantan province reported that PTB affected 7.3% of the total 65,075 births in 2023, amounting to 4,716 premature births. These statistics highlight the necessity for sustained efforts to improve maternal and child healthcare services, especially in regions like South Kalimantan with high PTB prevalence.

At Dr M. Ansari Saleh Hospital in Banjarmasin, the incidence of PTB has consistently risen over the past three years, with PTB cases increasing from 251 (8.34% of total deliveries) in 2020 to 285 cases (10.16%) in 2021 and 237 cases (13.07%) in 2022. This trend underscores the growing urgency for targeted interventions to mitigate PTB in the region. Initiatives such as the Childbirth Planning and Complication Prevention (P4K) program, integrated into the Maternal and Child Health (KIA) handbook, emphasize early detection and prevention of health risks during pregnancy, which is crucial for improving infant care quality (Bloom et al., 2019).

Recent studies have increasingly utilized machine learning (ML) algorithms, including Random Forest (RF) and Particle Swarm Optimization (PSO), to predict PTB risks. ML models leverage various pregnancy and childbirth features to achieve varying prediction accuracy. The Random Forest algorithm is particularly effective in handling complex, high-dimensional data, demonstrating a classification performance with a sensitivity of 91.4% and specificity of 99% for PTB prediction based on maternal characteristics and habits (Ngiam & Khor, 2019). Meanwhile, PSO has shown promise

in optimizing machine learning models by identifying the most relevant features for prediction, enhancing the accuracy of risk assessments.

This study employs machine learning techniques, specifically the HYBRID RF-PSO model, to address the clinical challenges of PTB prevention. The goal is to develop a predictive model capable of accurately identifying women at risk for PTB early in their pregnancy. Early detection can lead to timely interventions, improving mothers' and infants' health outcomes. Various machine learning algorithms will be evaluated based on their accuracy, sensitivity, and specificity, contributing to the growing body of knowledge on PTB risk prediction while utilizing medical records from pregnant women.

1.3 Problem Statement

Preterm birth (PTB) is a significant global public health issue, being a leading cause of death among children under five, particularly newborns, with approximately 15 million cases occurring annually worldwide. Effective prevention requires clinical management strategies, including close monitoring, early detection, and timely intervention. Promotive and preventive efforts in midwifery services, especially during pregnancy, are essential to reducing the risks associated with PTB. Identifying pregnant women at risk for PTB is crucial to preventing primary causes of infant mortality and morbidity. Despite the availability of various data sources, such as the Maternal and Child Health (KIA) book used by health workers, an effective prediction model for accurately forecasting PTB risk remains lacking. A robust predictive model employing machine

learning techniques, particularly Random Forest (RF) optimized through Particle Swarm Optimization (PSO), would enable health workers to interpret data better and implement preventive measures to mitigate PTB risk.

Although extensive research has been conducted in the health sector regarding PTB, a gap exists in designing predictive models for PTB during pregnancy, explicitly utilizing data from antenatal care services. Most existing studies focus on the intranasal or birth period, analyzing factors related to the mother, placenta, baby, and amniotic fluid using various laboratory tests and other supporting examinations. These studies primarily employ statistical analysis to examine PTB, often overlooking the critical pregnancy phase and applying machine learning algorithms, such as RF and PSO, to predict PTB during this crucial period.

The ability to predict PTB using social determinants, prenatal health characteristics, and sociodemographic factors could be instrumental in developing comprehensive health coverage, especially in low- and middle-income countries. In Indonesia, PTB remains a significant cause of infant mortality, contributing to 28.2% of all births, necessitating increased attention during pregnancy. In South Kalimantan, the incidence of PTB has been rising annually. One effort to address this has been the implementation of optimal antenatal services aimed at detecting high-risk pregnancies. However, these services have limited predictive power. PTB assessment in Indonesia, particularly in Banjarmasin, remains conventional, relying on diagnostic tools like the Poedji Rochjati Scorecard and screening data for early detection of high-risk pregnancies during the antenatal period.

While artificial intelligence (AI) and machine learning approaches have been increasingly employed globally to predict PTB based on various known risk factors, such research remains limited in Indonesia. The integration of machine learning algorithms, particularly RF optimized with PSO, into health services for predicting PTB has not been fully realized, and health institutions have yet to leverage these advanced methods as early prevention tools.

The findings of this study aim to address this gap by identifying key risk factors that can aid in predicting PTB, particularly in community services and suburban areas. The predictive model developed in this research is designed to support healthcare providers in identifying and warning pregnant women at risk of PTB, reducing diagnostic costs, minimizing complications, and ultimately preventing premature births. By utilizing data from the Maternal and Child Health (KIA) book maintained for all pregnant women in Indonesia, this research seeks to enhance the predictive capabilities of antenatal care services. By incorporating machine learning algorithms such as RF and optimizing them with PSO, this study contributes to the early prediction of PTB, enhancing the capacity for early intervention and improving maternal and child health outcomes.

1.4 Research Objectives

The study's objectives require further clarity to emphasize the expected contributions of this research. For example, is the primary focus on developing an innovative predictive model, or is it simply to design the predictive process? One critical question

that arises is why this research has chosen Particle Swarm Optimization (PSO) as the optimization technique to enhance the predictive model's performance. The background does not yet address the importance of PSO or its advantages over alternative methods.

Additionally, the second research objective is ambiguous: Is the study focused on model optimization or risk prediction? Although both objectives relate to the use of data and algorithms, the methodologies applied for optimization differ from those used for prediction. The rationale for using PSO should also include an argument explaining why this technique is optimal for PTB prediction.

The aims of this research are as follows:

1. To determine the most influential features in predicting preterm birth during pregnancy.
2. To design a process model for predicting preterm birth during pregnancy using Machine Learning algorithms enhanced with PSO.
3. To evaluate the effectiveness of the process model in predicting preterm birth during pregnancy using Machine Learning algorithms with PSO.

1.5 Research Questions

This research is specifically conducted to answer the following questions:

1. **RQ1**: What are the most impactful features in predicting preterm birth (PTB) during pregnancy?
2. **RQ2**: How can a process model be designed to predict PTB during pregnancy using machine learning algorithms with Particle Swarm Optimization (PSO)?
3. **RQ3**: How can the effectiveness of the process model for predicting PTB during pregnancy, using machine learning algorithms with PSO, be evaluated?

1.6 Significance of The Research

This research uses a machine learning algorithm to design a prediction model for preterm birth (PTB) during pregnancy. The significance of this research encompasses several vital points:

1. **Early Pregnancy Risk Prediction**: Preterm birth (PTB) is a significant health issue that can seriously impact babies born prematurely (Sari et al., 2024). Machine learning algorithms can assist in predicting the risk of PTB during pregnancy, enabling the identification of high-risk pregnant women. Early identification allows for appropriate preventive actions or interventions to reduce the likelihood of PTB (Solehati et al., 2020).

2. **Identification of Healthcare Services**: Machine learning algorithms facilitate the identification of healthcare services by analyzing various risk factors relevant to each individual. By understanding the unique characteristics of each pregnant woman, machine learning systems can provide tailored recommendations or alerts suited to their clinical needs (Włodarczyk et al., 2021).
3. **Improved Early Diagnosis**: Machine learning algorithms can analyze clinical information and relevant biomarkers to aid in the early detection of PTB risk. This enables medical professionals to proactively plan appropriate care and interventions, thus managing high-risk pregnancies more effectively (Park et al., 2021).
4. **Efficient Use of Resources**: More accurate PTB risk prediction allows for the efficient allocation of healthcare resources. This can reduce healthcare costs and ensure services are directed toward those needing them most (Berghella, 2022).
5. **Contribution to Future Research and Development**: Research on PTB prediction using machine learning may offer new insights into risk factors and underlying mechanisms. These findings could lead to developing new methods for preventing or treating PTB in the future.
6. **Enhanced Patient Outcomes**: Healthcare providers can implement timely interventions by accurately predicting PTB risk, potentially improving outcomes for both mothers and their infants.
7. **Integration of Technology in Maternal Healthcare**: The utilization of machine learning in PTB prediction exemplifies the integration of technology in maternal healthcare, highlighting the potential for innovative approaches to traditional health challenges.

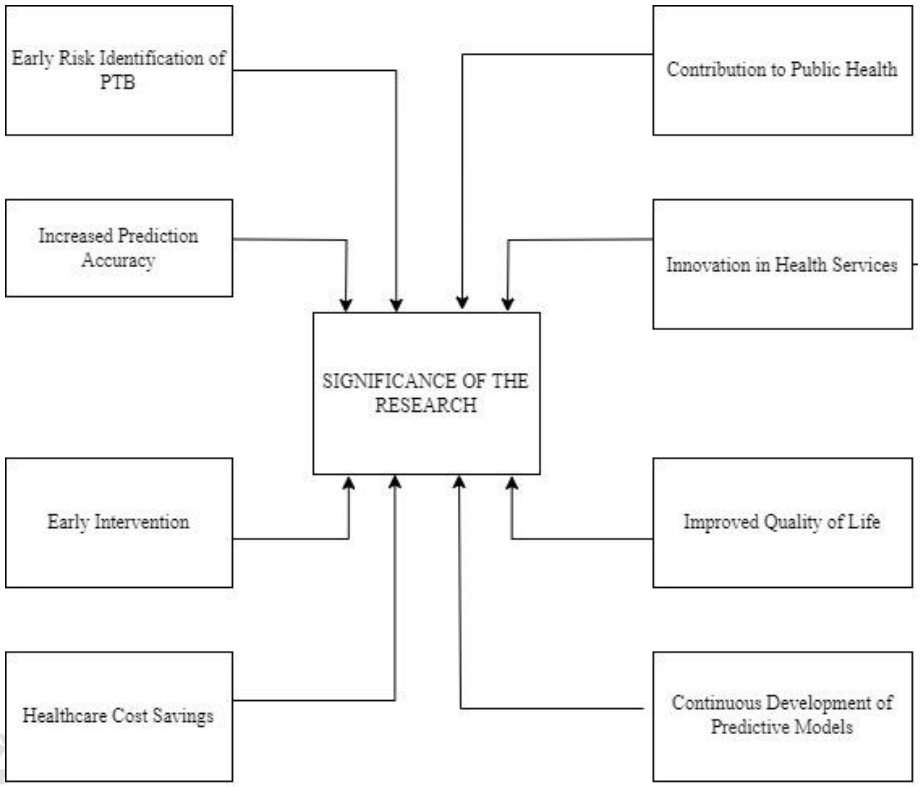
8. **Policy and Practice Implications**: The insights gained from this research can inform healthcare policies and practices, promoting evidence-based approaches to managing pregnancy and PTB risks.

Therefore, this research has the potential to significantly contribute to healthcare services for pregnant women and their unborn babies, paving the way for further advancements in the field of PTB prediction and prevention.

Eight key points highlight the significance of researching predicting preterm birth (PTB) during pregnancy using machine learning. First, early risk identification of PTB is facilitated by machine learning algorithms, which can identify important risk factors such as medical history, demographic characteristics, and biological biomarkers. This approach allows for personalized care for each pregnant woman, minimizing the risk of PTB through tailored interventions (Ramakrishnan, 2021). Additionally, machine learning provides increased prediction accuracy, offering more reliable risk assessments than traditional methods (Hsu et al., 2021).

Figure 1.1

Significance of the research



With improved accuracy, this research enables early intervention, allowing healthcare providers to act sooner with lifestyle changes, enhanced medical monitoring, or appropriate treatments to manage high-risk pregnancies (Nsugbe et al., 2021a). Early prediction and intervention also contribute to healthcare cost savings by reducing the long-term costs associated with treating and managing PTB. Through advanced data analysis, this research not only lowers the incidence of PTB but also provides new insights into the underlying factors, aiding in developing more effective prevention strategies (Berghella, 2022).

The findings of this study will further contribute to public health by informing clinical practices, public health policies, and educational programs for pregnant women about reducing PTB risks (Nsugbe et al., 2021b). Additionally, this research introduces innovation in health services, as machine learning technology leverages advanced data and analytics to transform maternal healthcare, offering more precise and efficient services (Podda et al., 2018).

Lastly, the potential outcomes of this research include improving the quality of life for pregnant women and their unborn babies by reducing the risks and complications associated with PTB (Rocha et al., 2021). Moreover, the study supports the continuous development of predictive models, as the results will provide a foundation for future models that can be updated and adapted to incorporate new data and evolving scientific discoveries (Rakesh Raja & Mukherjee, 2021).

Based on these points, it is evident that research into predicting PTB during pregnancy using machine learning has a significant impact on health services for pregnant women and unborn babies. It has the potential to enhance understanding and preventive measures against PTB in the future.

1.7 Scopes of The Research

The scope of this research is centred on predicting preterm birth (PTB) during pregnancy through various machine learning algorithms. The primary focus is identifying and analyzing the characteristics that influence the incidence of PTB by

leveraging structured data, such as medical records and maternal health information. Specifically, the research aims to determine critical factors—including medical history, demographic characteristics, and biological markers—contributing to PTB's likelihood. These factors will be identified and evaluated using machine learning techniques, drawing insights from previous obstetrics and maternal health studies.

This research is limited to predicting PTB using a set of well-established machine learning algorithms, including Random Forest (RF), K Nearest Neighbor (KNN), Decision Tree (DT), and Naive Bayes (NB). These algorithms were selected based on their performance in similar prediction tasks and their ability to handle complex datasets involving medical and clinical data. Each algorithm will be trained and tested to assess its accuracy and effectiveness in predicting PTB, allowing for comparing and evaluating its predictive capabilities.

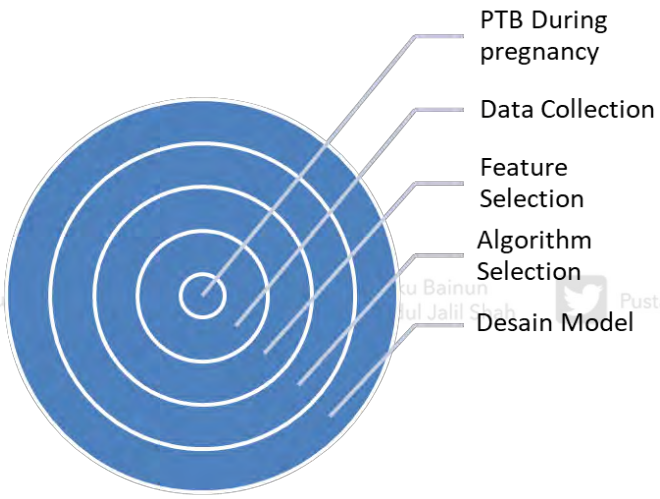
The dataset used in this study consists of medical records and Maternal Health Care (MHC) books from patients treated at Dr Ansari Saleh Hospital in Banjarmasin in 2022. These records provide a rich source of information, including details on prenatal checkups, maternal health conditions, and relevant clinical findings that may indicate PTB risk. By analyzing this dataset, the research aims to build predictive models that can accurately forecast the likelihood of PTB in pregnant women, facilitating early intervention and improving health outcomes.

As illustrated in Figure 1.2, the scope of this research encompasses the full spectrum of data collection, feature selection, and model development. It begins with data extraction from the hospital's medical records, preprocessing, and analysis of the

relevant risk factors. The machine learning algorithms are then applied to predict PTB, with the results evaluated and compared to establish the most effective model for clinical use. The findings of this study will contribute to ongoing efforts to enhance maternal healthcare by utilizing machine learning to anticipate and mitigate the risks of preterm birth.

Figure 1.2

Scope of the Research



The first step in the research process is data collection, which involves gathering comprehensive datasets that include maternal health records, demographic information, medical history, and lifestyle factors of pregnant women. These datasets may also encompass genetic markers, prenatal care visits, and any complications experienced during pregnancy. By collecting this extensive range of variables, the research ensures a holistic approach to identifying factors that may influence the likelihood of preterm birth (PTB).

Next, feature selection is crucial in narrowing down the most predictive variables from the collected data. This step involves performing statistical analyses and drawing on domain expertise to identify the most relevant factors for PTB prediction. The goal is to focus on features with the highest predictive power—such as specific medical conditions, biological markers, or socioeconomic factors—while minimizing noise from less relevant data.

The algorithm selection process begins after feature selection, where suitable machine learning algorithms are chosen for the prediction tasks. A comparison of several algorithms, including Naive Bayes (NB), Decision Tree (DT), Random Forest (RF), and K Nearest Neighbor (KNN), will be conducted. Additionally, techniques such as optimization with Particle Swarm Optimization (PSO) and cross-validation will enhance performance and model accuracy.

The next step is model design, which uses the selected algorithms and features to build predictive models. This phase includes data preprocessing, model training, and optimization to ensure that the models are accurate and reliable. Cross-validation techniques will be applied to validate the models' robustness and generalizability to new data. Finally, evaluation metrics will be defined to assess the performance of the predictive models. Metrics such as accuracy, precision, sensitivity, specificity, and ROC-AUC (Receiver Operating Characteristic - Area Under the Curve) will be utilized to evaluate how well the models predict PTB.

1.8 Importance of Research

Many predictions of preterm birth (PTB) are based on screening methods derived from health sector knowledge, utilizing statistical approaches to forecast the incidence of PTB during pregnancy. Consequently, precise and accurate predictions are essential. Machine learning (ML) can analyze nonlinear interactions within large datasets. Recent studies have demonstrated that ML technology can effectively characterize patient profiles by identifying diagnoses and prognoses for high-risk individuals. However, current methods predominantly focus on comparing the performance of various prediction techniques using datasets with differing attributes.

The research procedure encompasses several stages that lead to establishing a PTB risk prediction model based on statistical knowledge. This research is expected to contribute to:

1. Information Technology

This study aims to assist the health sector in integrating technology-based health services, mainly through machine learning (ML) applications for critical diagnostic decisions. By leveraging diverse data features to develop practical ML applications, new algorithms can achieve high accuracy, enhancing computerized midwifery services for pregnant women. Implementing ML can lead to more accurate predictions for PTB diagnosis during pregnancy, ultimately improving overall healthcare outcomes.

Integrating technology-based health services will facilitate advancements in clinical transformation within the health sector, particularly in midwifery.

2. Health

By equipping health workers with the tools to diagnose PTB during pregnancy, this research enables the implementation of promotive and preventive actions at the primary care level. This proactive approach anticipates PTB events through modern medical decision-making, necessitating accurate decision support with high sensitivity and specificity in predictive properties. Developing a clinically feasible predictive model will assist healthcare providers in effectively intervening and mitigating risks associated with PTB. Additionally, advocating for government policies aimed at reducing morbidity and mortality among infants and toddlers in Indonesia, particularly within midwifery and health services, is essential.

1.9 Operational Definition

Some terms are being used in this research that need to be comprehended according to these operational definitions to support the understanding of the research. This section defines essential terms as follows:

Table 1.1

Operational Definition for Research

Variable	Operational Definition
Maternal Mortality Rate (MMR)	Maternal death is the death of a woman during pregnancy, childbirth, and the puerperium due to her pregnancy
Infant Mortality Rate (IMR)	Infant mortality is the number of infants who die before reaching the age of 1 year.
Antenatal Care (ANC)	Visits of pregnant women to health services during pregnancy
Preterm Birth (PTB)	Premature Birth that occurs before the 37th week of pregnancy
Maternal Care Handbook (MCH)	The Maternal and Child Health Handbook is a guidebook for information and health records during pregnancy, childbirth, and children aged 6 years
Maternity Group	Maternity Cohort is a database to identify health problems during childbirth
Mother Characteristics	The characteristics are the characteristics possessed by pregnant women which are the characteristics
Obstetric History	Obstetric history is the result of a previous pregnancy or maternal and fetal complications, family history including hereditary disorders
Lifestyle	Lifestyle is part of the habits of pregnant women
Current Pregnancy	The current state of pregnancy is a condition experienced by pregnant women from conception to delivery
Public Health Center (PHC)	First-level healthcare facilities that prioritize services on promotive and preventive efforts
Maternal Age	Refers to the age of the mother when pregnant or giving birth.
Gestational Age	This is the gestational age calculated from the start of the last menstrual cycle.
Education	Refers to the mother's education level.

Variable	Operational Definition
Gravidity	The number of pregnancies a woman has had, including the current pregnancy
Interpregnacy interval	The time interval between the birth of the previous child and the conception of the next child.
History Abortion	The history of a woman who has had a previous incident where the pregnancy was miscarried or terminated.
History Sectio Caesaria	The history of a woman who has had a previous caesarean section for delivery.
Haemoglobin	This is a protein in the blood that transports oxygen from the lungs to the rest of the body.
Body Mass Indeks	A measurement that describes the proportion of a person's weight to their height. BMI is used to assess a person's nutritional status.
Machine Learning	Tools used to predict the incidence of preterm birth
Index Metrics	A measure or evaluation of model performance used to measure how well the model makes predictions on given data Some index metrics used in machine learning algorithms are: Accuracy, Precision, Sensitivity, Specificity, ROC Curve and AUC
Random Forest	Part of the machine learning Algorithm used to predict incidents of preterm birth
Naïve Bayes	Part of the machine learning Algorithm used to predict incidents of preterm birth
Decision Tree	Part of the machine learning Algorithm used to predict incidents of preterm birth
K Nearest Neighbor	Part of the machine learning Algorithm used to predict incidents of preterm birth
Particle Swarm Optimization (PSO)	Particle Swarm Optimization (PSO) parameters are used for optimization, namely by increasing the attribute weights of all the attributes or variables used, selecting attributes and selecting features, determining relevant attributes that have been used previously and eliminating irrelevant features that are not related to classification. .

Variable	Operational Definition
Cross - Validation	The technique tests model performance on previously unseen data, prevents overfitting, and selects the best model. This method allows efficient use of all data and helps in tuning model parameters.