RANDOM FOREST AND LOGISTIC REGRESSION ALGORITHMS PREDICTION MODEL IN VIOLENCE AGAINST WOMEN (VAW)

DINI RAHMAYANI

SULTAN IDRIS EDUCATION UNIVERSITY

2025

RANDOM FOREST AND LOGISTIC REGRESSION ALGORITHMS PREDICTION MODEL IN VIOLENCE AGAINST WOMEN (VAW)

DINI RAHMAYANI

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 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

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

UNIVERSITI PENDIDIKAN SULTAN IDRIS
SULTAN IDRIS EDUCATION UNIVERSITY

Please tick (√)	
Project Paper	
Masters by Research	
Master by Mixed Mode	
Doctor of Philosophy	√

INSTITUTE OF GRADUATE STUDIES

DECLARATION OF ORIGINAL WORK

This declaration is made on the 11 (day) March (month) 2025 (year)

i. Student's Declaration:

I, DINI RAHMAYANI, (MATRIC NO. P20211002735), Faculty of Computing and Meta-Technology hereby declare dissertation / thesis for Doctor of Philosophy in Artificial Intelligence titled RANDOM FOREST AND LOGISTIC REGRESSION ALGORITHMS PREDICTION MODEL IN VIOLENCE AGAINST WOMEN (VAW) 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. Madya Ts. Dr. Muhammad Modi Bin Lakulu hereby certify that the work entitled RANDOM FOREST AND LOGISTIC REGRESSION ALGORITHMS PREDICTION MODEL IN VIOLENCE AGAINST WOMEN (VAW) 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.

11/03/2025
Date

Signature of the Supervisor

UPSI/IPS-3/BO 31
Pind.: 01 m/s:1/1

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*: RANDOM FOREST AND LOGISTIC REGRESSION ALGORITHMS PREDICTION MODEL IN VIOLENCE AGAINST WOMEN (VAW)

No. Matrik /*Matric's No.*: P20211002735

Saya / *I* : DINI RAHMAYANI

(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 restircted information as specified by the organization where research was done.*
☑	**TIDAK TERHAD / *OPEN ACCESS***	

(Tandatangan Pelajar/ Signature)

Tarikh: 11 Mac 2025

Prof. Madya Ts Dr. Muhammad Modi Lakulu
Pensyarah
Fakulti Komputeran dan Meta-Teknologi
Universiti Pendidikan Sultan Idris
35900 Tanjong Malim, Perak

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

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 CONFIDENTAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction.

ACKNOWLEDGEMENT

The journey to this doctoral degree has been challenging yet rewarding, made possible by the support of many individuals. This thesis is submitted in partial fulfillment of the requirements for the degree of Doctor of Philosophy in Artificial Intelligence at the Sultan Idris Education University, Malaysia.

I am very grateful to my supervisor, Prof. Madya Ts. Dr. Muhammad Modi bin Lakulu, for her invaluable guidance, encouragement, and insights throughout this process. His extensive knowledge and high standards continue to inspire me to strive for excellence. To Assoc. Prof. Dr. Wang Shir Li, Prof. Madya Dr. Jafreezal Jaafar, and Prof. Madya Dr. Roslina binti Ibrahim, as evaluators for the suggestions given to improve this thesis, may Allah SWT reward them with goodness, Aamiin ya Robb.

I would like to express my sincere gratitude and deep appreciation to the following individuals. I express my sincere gratitude to Prof. Dr.-Ing. Maizatul Hayati binti Mohamad Yatim, Dean of the Faculty of Computing and Meta-Technology at the Sultan Idris Education University, for providing the opportunity and facilities necessary for this thesis. I am grateful to Dr. Suliana binti Sulaiman, Head of the Department of Software Engineering and Intelligent Technology at the Faculty of Computing and Meta-Technology, Sultan Idris Education University, for her facilitation, valuable inputs, and suggestions that improved this research.

I also express my gratitude to Yayasan Indah Banjarmasin and Universitas Sari Mulia who have given me the opportunity to pursue this doctoral education, may Universitas Sari Mulia continue to advance and develop.

On this occasion I would also like to express my gratitude to all parties who have greatly supported this research, namely the related agencies, including the Women's Empowerment and Child Protection Service of South Kalimantan Province, the South Kalimantan Provincial Police, especially the Women and Children Protection Unit Division, the Banjarmasin Religious Court who have been willing to facilitate the provision of data on violence against women in Banjarmasin City and in South Kalimantan Province, so that the data becomes very valuable information and contributes to this research. In addition, I would also like to thank the research location, namely the health service in the Banjarmasin City work area, so that this research can run smoothly. My deepest heart goes to my parents, beloved husband, daughters, and family that I love and care for, thank you for all the moral and material support that is immeasurable and very meaningful to me, may Allah SWT repay it with an appropriate reward, Amen, Robb. Finally, I express my deepest appreciation for the guidance and support from all parties that I have mentioned here. There are no words that can express my gratitude. I dedicate this work to all of you. Thank you.

ABSTRACT

Violence against women (VAW) is an urgent problem that not only has severe physical impacts but also psychological impacts. Predicting violence based on its type increases efficiency, accuracy, and early detection. This ensures that prevention, intervention, and policies can be more targeted and focused, as interventions for each type of violence vary and require a tailored approach. This study aims to develop the latest prediction model to predict VAW in each type which were physical, psychological, economic, and sexual violence. The methodology used is machine learning with the naïve bayes, random forest, and logistic regression algorithms because it provides a balance of efficiency, ease of interpretation, and the ability to handle complex data often encountered in the context of predicting VAW. A representative research sample of 600 married women, data collection through questionnaires designed to capture variables and characteristics associated with various types of VAW. The design of the prediction model in this study is very important for each type of violence because interventions for victims vary in their needs and challenges. Therefore, these findings have significant value in making precise and accurate predictions. Random forest had the best model with 98.75% accuracy, 99.4% AUC, 97.88% precision, 97.25% sensitivity, and 96.99% specificity. Furthermore, when considering VAW by type—physical violence, sexual violence, economic violence—the random forest algorithm performed best with accuracies of 93.33%, 90.00%, 85.83%, and the prediction of psychological violence, the logistic regression algorithm achieved the highest accuracy of 81.67%. This research contributes to improving predictive models for identifying potential risks of violence, developing datasets, and supporting policymakers with data-driven insights. These insights inform the design of a new model optimized to predict different types of VAW more accurately. Ultimately, this contributes to the development of more effective prevention and intervention strategies according to the type of violence.

MODEL RAMALAN ALGORITMA *RANDOM FOREST* DAN REGRESI LOGISTIK DALAM KEGANASAN TERHADAP WANITA

ABSTRAK

Keganasan terhadap wanita adalah masalah mendesak yang bukan sahaja memberi kesan fizikal yang teruk tetapi juga kesan psikologi. Meramalkan keganasan berdasarkan jenisnya meningkatkan kecekapan, ketepatan dan pengesanan awal, supaya pencegahan, campur tangan dan dasar boleh lebih disasarkan dan tertumpu, kerana campur tangan untuk setiap jenis keganasan berbeza-beza dan memerlukan pendekatan yang disesuaikan. Kajian ini bertujuan untuk membangunkan model ramalan terkini untuk meramalkan keganasan terhadap wanita dalam setiap jenis keganasan fizikal, psikologi, ekonomi dan seksual. Metodologi yang digunakan ialah pembelajaran mesin dengan Naïve Bayes, *Random Forest*, dan algoritma Regresi Logistik kerana ia memberikan keseimbangan kecekapan, kemudahan tafsiran dan keupayaan untuk mengendalikan data kompleks yang sering ditemui dalam konteks meramalkan keganasan terhadap wanita. Sampel penyelidikan yang mewakili 600 wanita berkahwin, pengumpulan data melalui soal selidik yang direka untuk mengenalpasti pembolehubah dan ciri yang dikaitkan dengan pelbagai jenis keganasan terhadap wanita. Oleh itu, penemuan ini mempunyai nilai yang signifikan dalam membuat ramalan yang tepat dan jitu. *Random forest* mempunyai model terbaik dengan ketepatan 98.75%, AUC 99.4%, ketepatan 97.88%, kepekaan 97.25% dan kekhususan 96.99%. Tambahan pula, apabila menganalisis keganasan terhadap wanita mengikut jenis, *Random Forest* menunjukkan prestasi unggul untuk keganasan fizikal, mencapai ketepatan 93.33%, dan untuk keganasan seksual dengan ketepatan 90.00%. Dalam kes keganasan ekonomi ini, algoritma *Random Forest* mencapai ketepatan 85.83%. Sebaliknya, untuk meramalkan keganasan psikologi, algoritma regresi logistik mencapai ketepatan tertinggi pada 81.67%. Penyelidikan ini menyumbang kepada penambahbaikan model ramalan yang boleh digunakan untuk meramalkan potensi risiko keganasan dan menyediakan pembangunan set data serta menyokong pembuat dasar dengan menyediakan cerapan terdorong data. Cerapan ini memaklumkan reka bentuk model baharu yang dioptimumkan untuk meramalkan pelbagai jenis keganasan terhadap wanita dengan lebih tepat. Akhirnya, ini menyumbang kepada pembangunan strategi pencegahan dan intervensi yang lebih berkesan mengikut jenis keganasan.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ACC	Accuracy
AI	Artificial Intelligence
AUC	Area Under Curve
DL	Deep Learning
GBV	Gender-Based Violence
LR	Logistic Regression
ML	Machine Learning
PRE-Precision	Precision
RF	Random Forest
SENS	Sensitivity
SPEC	Specificity
VAW	Violence Against Women
WHO	World Health Organization
VAC	Violence Against Women
RG	Regional Police
SB	South Borneo
STI	Sexual Transmitted Infection
Rehab	Rehabilitation
Soc.Serv.	Social Service

CPD	City Police Department
Res.Pol	Resor Police
RMW	Regional Minimum Wage
Tan.Bu	Tanah Bumbu
HSS	Hulu Sungai Selatan
HST	Hulu Sungai Tengah
HSU	Hulu Sungai Utara
Tala	Tanah Laut
Batola	Barito Kuala
CT	Case Total
CC	Case Closed

LIST OF APPENDICES

CHAPTER 1

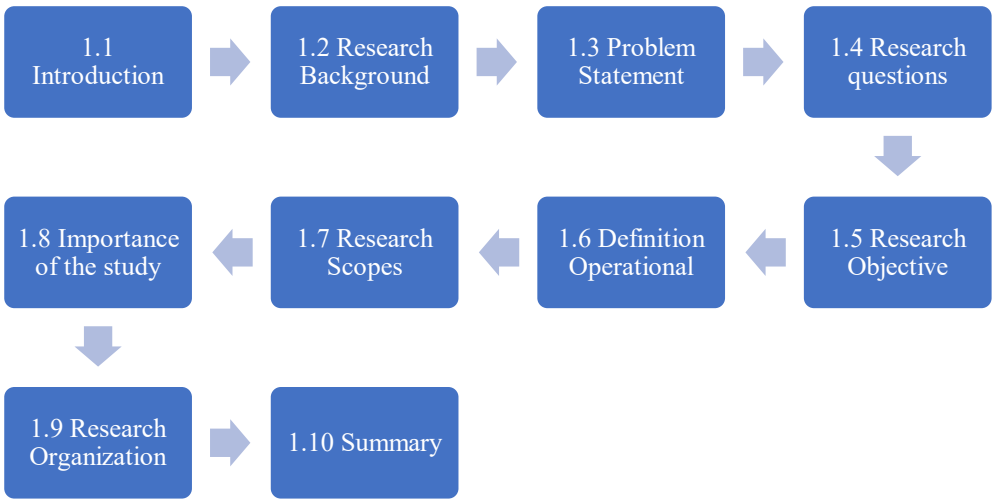
INTRODUCTION

1.1 Introduction

This chapter serves to introduce the purpose and concept of the research. Firstly, Section 1.2 delves into the Research Background. Secondly, Section 1.3 presents the problem statement as outlined by the researcher. Thirdly, Section 1.4 elaborates on the Research Questions. Following this, Section 1.5 outlines the Research Objectives. Subsequently, Section 1.6 provides Operational Definitions. This is followed by Section 1.7 defines the Research Scopes. Next, Section 1.8 which discusses the importance of the research. Then, Section 1.9 details the Research Organization. Lastly, Section 1.10 offers a summary of the research.

Figure 1.1

Introduction flow Diagram in Chapter one



1.2 Research Background

Artificial Intelligence (AI) has progressed remarkably across multiple sectors, showcasing its adaptability and potential for transformation. Recent developments in AI have brought it into fields like healthcare, agriculture, arts, manufacturing, and education, highlighting its role in enhancing efficiency, fostering creativity, and supporting decision-making. Recent progress in AI has greatly impacted healthcare, boosting diagnostic precision, personalized treatment options, and operational effectiveness. The adoption of AI in healthcare has been driven by the demand for innovative solutions, especially following the COVID-19 pandemic, which exposed the limitations of conventional healthcare systems and underscored AI's potential to overcome these issues (Alhasan & Hasaneen, 2021). In a similar vein, Wang and colleagues emphasize AI's contribution to improving diagnostic accuracy, particularly

through computer vision tools that can process imaging data more efficiently than conventional techniques (Wang et al, 2021).

Machine learning, a branch of AI, employs algorithms to process vast datasets, helping healthcare providers make well-informed decisions and enhance patient outcomes. This overview examines the latest advancements in AI and machine learning in tackling diverse healthcare challenges, emphasizing their uses, advantages, and the challenges they bring. One of the most impactful applications of machine learning in healthcare is in predictive analytics. By utilizing historical patient data, machine learning algorithms can forecast patient outcomes, which is essential for proactive healthcare management. Machine learning methods can be particularly useful for health services researchers, as they enable the prediction of various health outcomes based on complex datasets Doupe, Faghmous, and Basu (2019). This predictive capability allows healthcare providers to identify at-risk patients and implement timely interventions, thereby improving overall patient care.

The current health problem that continues to be a challenge in its resolution is violence against women. Violence against women remains a prevalent issue, resulting in both physical and psychological harm. Such violence can manifest anywhere, yet many women tend to respond passively to these acts and perceive them as normal, particularly when it occurs within the confines of the household. Women experiencing domestic violence often conceal such incidents due to familial shame, leading to a significant underreporting of cases—a phenomenon akin to an "Iceberg Phenomenon," where the number of unreported incidents far surpasses those that are known.

Women's rights are integral to achieving gender equality in global health objectives. Gender parity is a key focus area outlined in the Sustainable Development Goals (SDGs), specifically in Goal five among the 17 goals established to be accomplished by 2030. This goal aims to reduce the prevalence of violence against women and promote gender equality World Health Organization (WHO, 2018). According to the WHO's 2018 report, physical and sexual violence against women has reached epidemic proportions, impacting over a third of women worldwide. A recent study by the WHO underscores the pervasive nature of violence against women, affecting individuals across various socioeconomic strata. The study reveals that one in three women globally will experience physical or sexual violence at some point in their lives, often perpetrated by someone they know, such as a spouse, partner, family member, or friend. The report emphasizes that women of all ages, irrespective of their age, are vulnerable to such violence.

Claudia Garcia-Moreno highlights the severe health impacts of physical and sexual violence against women, supported by extensive research on the wide-ranging effects of gender-based violence (GBV). Recognized as a major public health concern, GBV affects over a billion people worldwide and has serious consequences for survivors' physical, reproductive, and mental well-being (Shrestha and Thakuri (2024). The connection between violence and women’s health is especially important, with studies showing that women subjected to intimate partner violence (IPV) face a higher risk of health issues such as unintended pregnancies, sexually transmitted infections, and mental health disorders like anxiety and depression (Costa et al., 2019).

In 2010, WHO, in collaboration with the London School of Hygiene & Tropical Medicine and the South African Medical Research Council, conducted a comprehensive data collection effort across 81 countries. The study revealed that Southeast Asia, the eastern Mediterranean region, and Africa were the area's most severely affected by violence against women, with approximately 37 percent of women experiencing violence from intimate partners. Furthermore, data on violence perpetrated by a combination of intimate partners and non-partner sexual violence indicated that Africa had the highest prevalence at 45.6 percent, followed by Southeast Asia at 40.2 percent. However, the report challenges the perception that violence is solely a problem in developing countries, highlighting that 32.7 percent of physical and sexual violence against women occurs in high-income countries (Muluneh, Stulz, Francis, & Agho, 2020).

According to the World Health Organization, (WHO,2021) domestic violence against women perpetrated by partners exhibits the highest prevalence rate, reaching 88%. This statistic corresponds to a total of 544,452 cases, with physical violence accounting for 49%, psychological violence for 23%, and sexual violence for 16%. In Indonesia, data on domestic violence against women inflicted by partners from 2018 to 2021 reveals a significant number of cases. In 2018, there were 8,234 reported cases, with physical violence accounting for 40%, sexual violence for 33%, and psychological violence for 9%. In 2019, the prevalence rose to 85%, with physical violence at 35%, sexual violence at 45%, and psychological violence at 5%. In 2020, the prevalence further increased to 97%, with physical violence accounting for 53% and sexual violence for 44%. In 2021, the prevalence remained high at 98%, with

79% of cases occurring within the domestic sphere (Siaran Pers Komnas Perempuan, 2022).

Women in Indonesia remain susceptible to becoming victims of violence. This circumstance was documented by (Siaran Pers Komnas Perempuan, 2022) in their report titled CATAHU, which provides insights into the dynamics of the number, variety, forms, domains, as well as structural, cultural, and legal obstacles in addressing GBV against women. The CATAHU 2022 report documents the evolving landscape of direct complaints received by (Siaran Pers Komnas Perempuan, 2022), service agencies, and Badilag. A total of 338,496 cases of GBV against women were documented, comprising 3,838 cases reported to (Siaran Pers Komnas Perempuan, 2022), 7,029 cases to service agencies, and 327,629 cases to Badilag.

These statistics reveal a significant 50% increase in GBV against women, totaling 338,496 cases in 2021 (up from 226,062 cases in 2020). There was a notable surge in cases reported to Badilag, with a 52% increase to 327,629 cases (from 215,694 in 2020). Complaints lodged with (Siaran Pers Komnas Perempuan, 2022) also saw a substantial rise of 80%, climbing from 2,134 cases in 2020 to 3,838 cases in 2021. Conversely, data from service institutions experienced a 15% decline, primarily attributed to certain institutions ceasing operations during the Covid-19 pandemic, as well as shortcomings in case documentation systems and limited resources. A preliminary study conducted on December 16, 2021, at the Office of Women's and Children's Empowerment in South Kalimantan revealed that out of 30 women who reported incidents, 21 had experienced domestic violence perpetrated by their partners.

Violence against women is prevalent in various regions of Indonesia, including the city of Banjarmasin in South Kalimantan Province. Banjarmasin, known as the "City of a Thousand Rivers," is characterized by its abundant waterways. The deeply ingrained culture within the South Kalimantan community, particularly in Banjarmasin, can either positively contribute to health or, conversely, serve as a negative influence. The prevailing culture significantly shapes the health behaviors of its inhabitants, influencing various societal phenomena, particularly concerning violence against women. This consideration formed the basis for selecting Banjarmasin as the research area.

Furthermore, the selection of the research location was informed by field data regarding incidents of violence against women. Both primary and secondary data were collected from various sources. Data on incidents of violence against women were sourced from the Women's Empowerment and Child Protection Service. It was observed that the number of reported cases of violence against women in South Kalimantan consistently increased over the years. For instance, in 2020, there were 17,574 cases, followed by 21,753 cases in 2021, 25,053 cases in 2022, and 26,161 cases in 2023. Notably, not all cases were reported to the police, as many were resolved through informal means. Moreover, there was a phenomenon where incidents of violence were considered a family disgrace, leading to their concealment and lack of reporting to the authorities.

Based on the data obtained, the South Kalimantan Regional Police handled 155 cases of violence against women in 2020, 129 cases in 2021, and 220 cases in 2023. This indicates an upward trend in the number of cases addressed by the South

Kalimantan Regional Police over the years. However, it is worth noting that there is a known phenomenon, as reported by the police, wherein many cases of violence are reported but are not pursued further due to various factors, including the preference for resolving the matter amicably by the reporting party, among others.

In addition to police data, researchers also gathered information from various other sources, including the Banjarmasin Religious Court. One notable impact of violence against women is the rise in divorce rates, particularly attributed to instances of domestic violence. According to data from the Banjarmasin Religious Court, the number of divorce cases due to domestic violence has fluctuated over the years. For instance, there were 460 cases in 2020, 581 cases in 2021, 136 cases in 2022, and 93 cases in 2023. Although there was a sharp increase in divorce rates in 2021 followed by a decrease in the subsequent years, the issue of divorce stemming from violence against women remains a significant concern for various stakeholders. It is imperative that immediate attention be given to resolving this issue to address the ongoing problem of violence against women in Banjarmasin, South Kalimantan.

Violence against women represents a global public health crisis of significant magnitude. It is estimated that one in every three women worldwide has experienced physical, emotional, or sexual violence. Alarmingly, a substantial portion of this violence remains unreported. The prevalence of Violence Against Women (VAW) has sparked increasing interest in addressing these issues from the perspective of computer science (CS) and engineering (Rodríguez et al., 2021). However, there are still gaps in predictive models for accurately classifying the types of violence using Machine Learning techniques. Such models are urgently needed for swift and precise

intervention (Hossain, Najib, & Islam, 2020). Additionally, there remains a lack of comprehensive approaches for assessing the scale of domestic violence among women using Machine Learning methods. This is crucial as determining the severity of domestic violence, classified from mild to moderate, informs subsequent management strategies (McDougal et al., 2021); (Saboya, Angel, & Loaiza, 2019). Moreover, detailed descriptions of domestic violence behaviors are lacking, hindering classification into specific types of domestic violence (Gutiérrez, Vallejo, & Hernandez, 2019).

While the available data may be minimal, it does not reflect the true scale of violence against women. Many cases go unreported, contributing to what is known as the Iceberg Phenomenon. A preliminary study conducted on health services in 2022 revealed several findings: identification of violence against women relies solely on community complaints, primarily addressing physical violence while neglecting psychological and economic abuse; domestic violence interventions mainly focus on physical violence with no provisions for psychological and economic aspects; there is a lack of initial instruments for identifying violence against women, and predictive measures rely solely on manual methods without technological integration.

The existing gap poses a significant risk of increased incidents of violence against women, particularly in Indonesia. Addressing this gap is crucial to effectively combating the issue of violence against women. Accurate and timely prediction of violence incidents is essential to facilitate prompt and appropriate interventions. Current approaches to addressing this issue involve collaborative efforts across various scientific fields, particularly leveraging artificial intelligence

solutions in the digital era. This approach, as suggested by several articles, aims to tackle health issues, especially women's reproductive health, through interdisciplinary collaboration and the application of artificial intelligence methodologies.

The use of artificial intelligence (AI) through machine learning presents an effective approach to predict incidents of violence against women with high accuracy. The application of machine learning to predict domestic violence in Bangladesh during the COVID-19 outbreak, demonstrating the potential of machine learning-based models to predict domestic violence not only in Bangladesh but also in other regions globally (Hossain et al., 2021). Given the contextual background of the problem, the application of machine learning proved to be very suitable for predicting violence against women. However, in this study, the prediction of violence against women that was built was not only limited to predicting violence against women in general, but also developed a prediction model into a new prediction model technique that focuses on predicting models for types of violence against women by applying several algorithms to machine learning. Fast and accurate predictions facilitated by machine learning enable comprehensive and rapid handling of violence against women. Analysis of various articles underlines the efficacy of machine learning in making predictions quickly and accurately in this context.

1.3 Problem Statement

Prevalence of Violence Against Women (VAW) has spurred increasing interest in addressing these issues from the perspectives of computer science (CS) and engineering (Rodríguez, Diaz, iranda, Trujillo, & Mejia, 2021). The application of machine learning algorithms, including random forest, naive Bayes, and logistic regression, to predict violence against women presents significant opportunities and challenges.

The challenges and issues encountered in the use of these algorithms include that the effectiveness in predicting violence against women can be affected by various factors, including data quality, interpretability, and ethical considerations, while the "black box" nature of the random forest model can make it difficult for practitioners to understand the underlying reasoning behind certain predictions, and the use of naive Bayes assumes independence among predictors, thereby simplifying the modeling process and making it computationally efficient. However, the independence assumption may not hold in real-world scenarios, especially in the context of domestic violence, where multiple factors are often interrelated (Coll et al., 2021); (Sumra et al., 2023). Furthermore, the logistic regression algorithm is a widely used statistical method for modeling binary outcomes, making it well-suited for predicting the likelihood of violence against women. However, logistic regression may not perform well when dealing with highly complex data sets or when the relationships between variables are nonlinear, which may limit its effectiveness in certain contexts (Chernet & Cherie, 2020). Given these issues, addressing the

challenges of data bias, interpretability, and ethical considerations are highly considered in this study to realize the responsible application of this technology.

Based on the problems and answering these challenges, this study develops the latest model of predicting violence against women by considering the possibility of data bias, interpretability, and ethical considerations with one of its efforts using a data set sourced from primary data, in addition this study also improves data quality by identifying the importance features of violence against women by utilizing these characteristics in designing a predictive model of violence against women and evaluating them, which aims to optimize performance. In addition, in the health sector, these characteristics will be a reference for formulating programs by the Banjarmasin City Health Office to prevent and intervene in incidents of violence against women.

1.4 Research Questions

RQ1 : How to conduct data quality assessment through important features to the violence against women?

RQ2 : How to develop new predictive models for violence against women and its types?

RQ3 : How to evaluate the performance of predictive models of violence against women and its types based on accuracy, AUC, specificity, and sensitivity?

Table 1.1

The Relationship of research objective and research question

Research Objective	Research Questions
To conduct data quality assessment through important features to the violence against women.	How to conduct data quality assessment through important features to the violence against women?
To develop new prediction model for violence against women and its types.	How to develop new predictive models for violence against women and its types?
To evaluate the performance of predictive models of violence against women and its types based on accuracy, AUC, specificity, and sensitivity.	How to evaluate the performance of predictive models of violence against women and its types based on accuracy, AUC, specificity, and sensitivity?

1.5 Research Objectives

This study aims to:

1. To conduct data quality assessment through important features to the violence against women.
2. To develop new prediction model for violence against women and its type.
3. To evaluate the performance of predictive models of violence against women and its types based on accuracy, AUC, specificity, and sensitivity.

1.6 Operational Definition

This section elaborates on the significance of understanding the following terms:

1. Reproductive health: Individual health encompasses physical, mental, and social well-being in all aspects related to the reproductive system. It is not merely the absence of disease or dysfunction but also includes a positive and respectful approach to sexual health, free from coercion and violence (Roudsari, Sharifi, & Goudarzi, 2023).
2. Women's reproductive health: Overall well-being, encompassing physical, mental, and social health related to the female reproductive system, is closely tied to autonomy, decision-making power, and access to reproductive health services such as family planning and Sexual Transmitted Infection (STI) prevention. This emphasizes the significance of overall health and the quality of women's lives (Kalidhasan & Arumugam, 2020); (Ramudu & Balaprasanna, 2023); (WHO, 2021).
3. Violence Against Women: a range of actions designed to instill fear, assert control, and undermine women's rights and dignity, manifesting in various forms such as physical, psychological, financial, and sexual harassment (Öhman, Burman, Carbin, & Edin, 2020); (Yari, Zahednezhad, Gheslagh, & Kurdi, 2021).
4. Physical violence: A type of violence characterized by the intentional application of physical force or power against oneself, another individual, or a group, leading to injury, death, psychological damage, developmental issues, or deprivation (Harris, Conklin, Woods, & Cortes, 2023).
5. Psychological violence: A type of violence that includes behaviors intended to inflict harm through non-physical means, such as intimidation, isolation, verbal

abuse, victim blaming, and controlling daily activities (Can & Hayli, 2020); (O'Mullan, Mainey, Greer, & Breen, 2022); (Zerihun, Tesfaye, Deyessa, & Bekele, 2021).

6. Sexual violence: A form of violence characterized by actions or behaviors that target a person's sexuality or sexual organs without consent, often involving force, threats, or coercion (Escura, Angles, Ribera, & Castelo, 2022); (Mekonnen, Lakew, & Melese, 2022).
7. Economic violence: A type of violence that includes controlling, exploiting, or depriving a person of their financial resources or economic independence (Alsawalqa, 2020); (Bonamigo, Carvalho, & Clubas, 2021).

1.7 Research Scope

This research focuses exclusively on predicting occurrences of violence against women utilizing Machine Learning techniques. The dataset utilized in this study comprises information on violence incidents involving married women in Banjarmasin. It is important to note that this study does not encompass violence against children or men.

1.8 Importance of the Research

Given the background outlined above, the prevalence of violence against women remains notably high in Indonesia, particularly in Banjarmasin. Hence, it becomes

imperative to implement preventive measures aimed at reducing such violence. Consequently, the outcomes of this research can contribute to:

1. To Artificial Intelligence

a. Propose and advocate the development of Artificial Intelligence applications through machine learning with the best algorithm based on the results of this study, namely a new prediction model using random forest and logistic regression algorithms as an early identification of violence against women according to the types of violence.

b. Utilizing Artificial Intelligence through machine learning using the random forest algorithm as the best algorithm to predict violence against women effectively, quickly, and accurately so that it becomes the right solution and makes it easier to carry out timely interventions.

2. To Health

Understanding the characteristics and classification of violence against women through machine learning methods is crucial to establish Indonesian government policies, particularly in Banjarmasin, by understanding the characteristics, in this case the characteristics of women are comprised of features that can be used as interventions in planning health programs in health services, if we understand the characteristics that can contribute to the occurrence of violence, we can intervene early by formulating a health program policy so that the health program that is implemented will remain targeted according to the characteristics of these women. Utilizing the most accurate-performing method as a tool to predict

incidents of violence against women can aid in policy formulation and implementation.

Predictions of violence types can also serve in evaluating the effectiveness of existing programs and policies. By monitoring and analyzing data on the occurrence of different types of violence, it becomes possible to assess the efficacy of interventions in reducing violence and providing support to victims.

1.9 Research Organization.

This thesis comprises five distinct chapters. The first chapter provides an overview of Violence Against Women, including the study's background, problem statement, research questions, objectives, scope, and importance of the research. The second chapter presents a review of relevant literature. The third chapter outlines the research methodology employed in this study. The fourth chapter presents data analyses and the findings of the best framework. Finally, the fifth chapter delves into the discussions, conclusions, recommendations, limitations of the study, and suggestions for further research.

Research organization included in each chapter is a scheme of the things or topics discussed in each chapter, to make it easier to understand the things that will be discussed in the chapter.

1.10 Summary

This chapter delves into the background of the issue and the rationale behind its significance. The research topic focuses on violence against women, a public health concern often likened to an iceberg phenomenon where a significant portion of data remains undisclosed compared to what is known. Violence against women poses an urgent challenge, impacting women not only physically but also inflicting severe psychological trauma. Early identification of such instances is crucial. Hence, this study proposes predicting the occurrence of violence against women through a machine learning approach. Integrating health sciences with artificial intelligence, particularly through machine learning applications, offers a more effective and comprehensive solution to address current health challenges.