

**Universiti Teknologi MARA**

**Sentiment Analysis of Domestic  
Violence Prediction Using Naive Bayes  
Algorithm**

**Nurul Izzah Binti Mohd Rahiman**

**Thesis submitted in fulfilment of the requirements  
for Bachelor of Computer Science (Hons.) College  
of Computing, Informatics, and Mathematics**

**January 2024**

## **ACKNOWLEDGEMENT**

First and foremost, I start my gratitude for His Almighty for giving me the blessing of a good health that allows me to complete this research successfully within the given time. Next, to my advisor, Ts. Dr. Syarifah Adilah Mohamed Yusoff, who accepted me to be supervised under her guidance. She gives her utmost on guiding me on finishing this project. Her support, comments, and suggestions sure helped me a lot on understanding my understanding regarding the project. I appreciated her kindness and patience and respected her diligence on helping me doing all kind research throughout this project.

Apart from that, I extend my gratitude to Madam Ummu Fatimah Binti Mohd Bahrin, who is my CSP600 and CSP650 lecturer, for her insights and knowledge-sharing for my project. I am thankful for everything that she has shared and all the advice she gave while I'm having problems with this project. Not to forget my beloved parents and sister, being my biggest supporter and shared their opinion and guidance on my research.

Finally, I want to give my appreciation towards my batchmates, especially my classmates, for their support and advice given. I appreciate and thank them for being my emotional support while finishing my project. Without them that I've mentioned, I probably will be having such a hard time on completing this project.

## ABSTRACT

This research delves into the widespread issue of domestic violence, emphasizing its severe impact on individuals and society globally. The surge in domestic violence during the COVID-19 pandemic, as highlighted by UN Women's survey, particularly in countries like Kenya, sets the stage for the research problem. Recognizing the lack of public awareness and understanding of attitudes towards domestic violence, the study proposes using sentiment analysis on Twitter data to monitor real-time public sentiment. The research objectives focus on studying and applying the Naïve Bayes algorithm for sentiment analysis on tweets related to domestic violence, aiming to provide insights for researchers, government agencies, policymakers, and the public and develop a prediction model using Naïve Bayes algorithm to evaluate its performance. The scope involves using English language tweets collected from March 2021 to November 2023, limiting the data to the topic of domestic violence. Few Naïve Bayes classifiers are used to compare the accuracy of the Naïve Bayes algorithm and parameter tuning also done on the classifiers. Resampling is used to handle the imbalance dataset. This research also compares using VADER and SentiWordNet lexicon to compare which has the best accuracy. The evaluation of algorithms consists of comparing the accuracy, specificity, and other evaluation metrics. Based on the results, Bernoulli classifier has the best accuracy of 94% while Multinomial has an accuracy of 93%. The best ratio of data to be used are 80:20 with VADER lexicon approach.

**Keywords:** domestic violence, sentiment analysis, prediction model, Naïve Bayes algorithm, Multinomial classifier, Bernoulli classifier, accuracy, specificity, evaluation metrics, ratio, lexicon approach, VADER, SentiWordNet

# TABLE OF CONTENTS

<b>CONTENTS</b>	<b>PAGE</b>
<b>SUPERVISOR APPROVAL</b> .....	iii
<b>STUDENT DECLARATION</b> .....	iv
<b>ACKNOWLEDGEMENT</b> .....	v
<b>ABSTRACT</b> .....	vi
<b>TABLE OF CONTENTS</b> .....	vii
<b>LIST OF TABLES</b> .....	x
<b>LIST OF FIGURES</b> .....	xi
<b>LIST OF ABBREVIATIONS</b> .....	xiv

## **CHAPTER ONE: INTRODUCTION**

1.1. Background of Study .....	1
1.2. Problem Statement.....	3
1.3. Research Objective .....	5
1.4. Research Scope.....	5
1.5. Research Significance.....	6
1.6. Overview of Research Framework .....	7
1.7. Summary.....	8

## **CHAPTER TWO: LITERATURE REVIEW**

2.1. Domestic Violence .....	9
2.1.1. What is Domestic Violence.....	9
2.1.2. Cause of Domestic Violence .....	10
2.1.3. How Domestic Violence Affects People .....	11

2.2.	Sentiment Analysis .....	12
2.2.1.	Level of Sentiment Analysis .....	12
2.2.2.	Approach Used in Sentiment Analysis .....	14
2.3.	Naïve Bayes Algorithm .....	17
2.3.1.	Introduction to Naïve Bayes .....	17
2.3.2.	Details Mechanism and Classification of Naïve Bayes .....	18
2.3.3.	Advantages and Disadvantages of Naïve Bayes .....	20
2.4.	Implementation of Naïve Bayes Algorithm in Various Problem .....	23
2.5.	Similar Works .....	29
2.6.	The Implication of Literature Review .....	35
2.7.	Summary .....	36

## **CHAPTER THREE: METHODOLOGY**

3.1	Overview of Research Methodology Framework .....	37
3.1.1.	Details of Research Framework .....	40
3.2	Preliminary Study .....	41
3.2.1.	Literature Review .....	41
3.2.2.	Knowledge Acquisition .....	42
3.2.3.	Data Collection .....	43
3.3	Design and Implementation .....	47
3.3.1.	Research Framework .....	47
3.3.2.	Dashboard Design Interface .....	51
3.3.3.	Dashboard Development .....	51
3.3.4.	Data Preparation .....	52
3.3.5.	Applying Naive Bayes Algorithm .....	54
3.4	Evaluation Phase .....	55
3.5	Gantt Chart .....	57
3.6	Summary .....	58