

## Contents

<b>APPROVAL PAGE</b>	<b>i</b>
<b>PLAGIARISM STATEMENT</b>	<b>ii</b>
<b>MOTTO</b>	<b>iii</b>
<b>DEDICATION</b>	<b>iv</b>
<b>FOREWORD</b>	<b>v</b>
<b>TABLE OF CONTENTS</b>	<b>ix</b>
<b>LIST OF TABLES</b>	<b>x</b>
<b>LIST OF FIGURES</b>	<b>xi</b>
<b>ABSTRAK</b>	<b>xiii</b>
<b>ABSTRACT</b>	<b>xiv</b>
<b>1 Introduction</b>	<b>1</b>
1.1 Background . . . . .	1
1.2 Research Problem . . . . .	3
1.3 Research Scope . . . . .	4
1.4 Research Objectives . . . . .	4
1.5 Research Advantages . . . . .	5
<b>2 Literature Review</b>	<b>6</b>
<b>3 Theory</b>	<b>12</b>
3.1 Clickbait . . . . .	12
3.2 Clickbait Detection . . . . .	13
3.3 Hybrid Approach . . . . .	13
3.4 Preprocessing . . . . .	13
3.4.1 Case Folding . . . . .	14
3.4.2 Tokenization . . . . .	14
3.5 Feature Extraction . . . . .	14

3.5.1	Word Embedding . . . . .	15
3.6	Convolutional Neural Network . . . . .	17
3.6.1	Convolutional Layer . . . . .	18
3.6.2	Pooling Layer . . . . .	18
3.7	Logistic Regression . . . . .	18
3.8	Support Vector Machine . . . . .	19
3.9	Performance Evaluation . . . . .	20
3.9.1	Accuracy . . . . .	21
3.9.2	Recall . . . . .	22
3.9.3	Precision . . . . .	22
3.9.4	F1-Score . . . . .	22
<b>4</b>	<b>Research Methodology</b>	<b>23</b>
4.1	Research Description . . . . .	23
4.2	Tools and Materials . . . . .	24
4.2.1	Tools . . . . .	24
4.2.2	Materials . . . . .	24
4.3	Research Phase . . . . .	24
4.4	System Design . . . . .	25
4.5	Dataset . . . . .	26
4.6	Preprocessing . . . . .	26
4.6.1	Case Folding . . . . .	28
4.6.2	Tokenization . . . . .	28
4.6.3	Data splitting . . . . .	28
4.7	Word Embedding . . . . .	29
4.8	Traditional Machine Learning . . . . .	29
4.8.1	Logistic Regression . . . . .	30
4.8.2	Support Vector Machine . . . . .	30
4.9	Deep Learning . . . . .	31
4.10	Hybrid . . . . .	32
4.11	Performance Evaluation . . . . .	34
<b>5</b>	<b>Implementation</b>	<b>36</b>
5.1	Dataset . . . . .	36
5.1.1	Dataset Merging . . . . .	37
5.2	Preprocessing . . . . .	39
5.2.1	Case Folding . . . . .	39
5.2.2	Punctuation and Numerical Removal . . . . .	39
5.2.3	Word Stemming . . . . .	40
5.2.4	Tokenization . . . . .	41
5.3	Word Embedding . . . . .	42
5.4	Model Approach . . . . .	43
5.4.1	Traditional Machine Learning Approach . . . . .	44
5.4.2	Convolutional Neural Network . . . . .	46
5.4.3	Hybrid Approach . . . . .	49

5.5	Model Training . . . . .	49
<b>6</b>	<b>Results and Discussion</b>	<b>52</b>
6.1	Model Performances . . . . .	52
6.1.1	Traditional Machine Learning . . . . .	52
6.1.2	Deep Learning . . . . .	54
6.1.3	Hybrid Approach . . . . .	54
6.2	Discussion . . . . .	55
6.2.1	Traditional Machine Learning Approach . . . . .	55
6.2.2	Deep Learning . . . . .	55
6.2.3	Hybrid Approach . . . . .	57
<b>7</b>	<b>Conclusion and Future Works</b>	<b>59</b>
7.1	Conclusion . . . . .	59
7.2	Future Works . . . . .	60