

Contents

APPROVAL PAGE	i
PLAGIARISM STATEMENT	ii
MOTTO	iii
DEDICATION	iv
FOREWORDS	v
TABLE OF CONTENTS	x
LIST OF TABLES	xi
LIST OF FIGURES	xii
ABSTRAK	xiv
ABSTRACT	xv
1 CHAPTER I INTRODUCTION	1
1.1 Background	1
1.2 Research Problem	4
1.3 Research Questions	4
1.4 Research Scope	5
1.5 Research Objective	5
1.6 Research Advantages	5
1.7 Research Description	6
1.8 Research Organization	7
2 CHAPTER II LITERATURE REVIEW	8
3 CHAPTER III THEORITICAL BASIS	16
3.1 Clickbait	16
3.2 Style	16
3.3 Text Style Transfer	17
3.4 Natural Language Generation	17
3.5 Text Preprocessing	18
3.6 Case Folding	18
3.7 Tokenization	19
3.8 Word Embedding	19
3.9 Prototype Editing	20

3.10	N-grams Language Models	21
3.11	Term Frequency-Inverse Document Frequency Weight Overlap	21
3.12	Salience	23
3.13	Naive Bayes Classifier	23
3.14	Encoder-Decoder Model	24
3.15	Neural Network	25
3.16	Neural Machine Translation	25
3.17	Recurrent Neural Network	25
3.18	Gated Recurrent Unit	26
3.19	Beam Search	28
3.20	Bilingual Evaluation Understudy	28
4	RESEARCH METHODOLOGY	30
4.1	Research Phases	30
4.2	Tools and Datasets	31
4.3	Text Preprocessing	32
4.3.1	Case folding Tokenization	32
4.3.2	Whitespace Tokenization	32
4.4	Feature Extraction	32
4.4.1	Content and Attribute Extraction	33
4.4.2	Similar Text Extraction	35
4.5	Datasets Preprocessing	36
4.5.1	Vocabulary Dictionary Generation	37
4.5.2	Pairs Generation	37
4.5.3	Datasets Split	37
4.6	Model Preparation	38
4.6.1	Word to Index Conversion	39
4.6.2	Index to Embedding Conversion	39
4.7	Model Learning	39
4.7.1	Encoder Decoder Model	40
4.7.2	Training Process	41
4.7.3	Evaluation Process	42
4.8	Text Generation	43
5	CHAPTER V IMPLEMENTATION	45
5.1	Tools and Datasets	45
5.2	Text Preprocessing	48
5.2.1	Casefolding	48
5.2.2	Whitespace Tokenization	49
5.3	Feature Extraction	50
5.3.1	Content and Attribute Extraction	50
5.3.2	Similar Text Extraction	54
5.4	Datasets Preprocessing	59
5.4.1	Vocabulary Dictionary Generation	59
5.4.2	Pairs Generation	59

5.4.3	Datasets Split	60
5.5	Model Preparation	60
5.5.1	Word to Index Conversion	62
5.5.2	Index to Embedding Conversion	62
5.6	Model Learning	63
5.6.1	Encoder Decoder Model	64
5.6.2	Training Process	65
5.6.3	Evaluation Process	72
5.7	Text Generation	75
6	CHAPTER VI RESULT AND DISCUSSION	81
6.1	Datasets Analysis	81
6.2	Model Performance Analysis	81
6.2.1	DeleteOnly Approach	82
6.2.2	DeleteRetrive Approach	84
6.3	Performance Metrics Comparison	88
7	CHAPTER VII CONCLUSION AND FUTURE WORKS	91
7.1	Conclusion	91
7.2	Future of Works	92