

TABLE OF CONTENT

HALAMAN PENGESAHAN	i
STATEMENT	ii
FOREWORD	iv
TABLE OF CONTENT	v
LIST OF FIGURES	vii
LIST OF TABLES	viii
ABSTRACT	ix
ABSTRAK	x
CHAPTER I INTRODUCTION	1
1.1 Research Background	1
1.2 Research Problem	3
1.3 Research Scope	3
1.4 Research Objective	3
1.5 Research Benefits	4
1.6 Thesis Organization	4
CHAPTER II LITERATURE REVIEW	6
CHAPTER III BASIC THEORY	10
3.1 Clickbait Detection	10
3.2 Self-training	10
3.3 Support Vector Machine	11
3.4 K-means Clustering	13
3.5 Fasttext	15
3.6 Term Frequency – Inverse Document Frequency (TF-IDF)	16
3.7 Confusion Matrix	17
3.8 Precision	18
3.9 Recall	18
3.10 F-Measure	18
3.11 Accuracy	18
3.12 Threshold-surpassing Labelled Data	19
CHAPTER IV RESEARCH METHODOLOGY	20
4.1 Research Description	20
4.2 Tools and Materials	20
4.3 Research Phase	20
4.3.1 Literature Study	20
4.3.2 System Design	21

4.3.3	Data Acquisition	22
4.3.4	Data Pre-processing	24
4.3.5	Word Representation	25
4.3.6	K-means Clustering	25
4.3.7	Self-training	26
4.3.8	Performance Evaluation.....	29
CHAPTER V IMPLEMENTATION.....		31
5.1	Specification of Hardware and Software	31
5.2	Data Pre-processing Implementation	31
5.2.1	Stemming	32
5.2.2	Tokenization	32
5.3	Word Representation Implementation	32
5.3.1	Fasttext.....	32
5.3.2	Term Frequency – Inverse Document Frequency.....	33
5.4	Semi-supervised Learning Implementation	34
5.4.1	Traditional Self-Learning Method.....	34
5.4.2	Modified Self-Training using K-Means Clustering.....	35
5.5	Model Evaluation	36
CHAPTER VI RESULT AND DISCUSSION		38
6.1	Supervised Learning Model	38
6.2	Semi Supervised Learning Model Training and Tuning Result.....	38
6.2.1	Traditional Self-Training Model.....	39
6.2.2	Modified Self-Training Model	41
6.3	Model Performance Evaluation Result	45
6.3.1	Model Performance Evaluation using TF-IDF	46
6.3.2	Model Performance Evaluation using Fasttext.....	47
6.4	Final Analysis.....	48
CHAPTER VII CONCLUSION.....		50
7.1	Conclusion.....	50
7.2	Suggestions.....	50
REFERENCES		52