

TABLE OF CONTENTS

Semi-supervised Learning Method for Clickbait Detection in Bahasa Indonesia	i
Semi-supervised Learning Method for Clickbait Detection in Bahasa Indonesia	ii
APPROVAL PAGE	iii
STATEMENT	v
PREFACE	vii
TABLE OF CONTENTS	viii
TABLE OF TABLES	x
TABLE OF FIGURES	xi
Abstract	xii
CHAPTER I INTRODUCTION	1
1.1 Research Background	1
1.2 Research Problem	4
1.3 Research Scope	4
1.4 Research Objective	4
1.5 Research Benefit	5
CHAPTER II LITERATURE REVIEW	6
CHAPTER III THEORETICAL BASIS	18
3.1 Text Representation	18
3.1.1 Term Frequency-Inverse Document Frequency (TF-IDF)	18
3.2 Semi-supervised Learning	19
3.2.1 Self-training	20
3.2.2 Co-training	21
3.3 Supervised learning	22
3.3.1 Logistic regression	22
3.3.2 Support vector machine	23
3.4 Performance Evaluation	24
3.4.1 Accuracy	24
3.4.2 Precision	25
3.4.3 Recall	25
3.4.4 F1-Score	25

CHAPTER IV RESEARCH METHODOLOGY	26
4.1 Research Description	26
4.2 Tools and Material	26
4.2.1 Tools	26
4.2.2 Materials	26
4.3 Research Phases	27
4.4 System Design	28
4.5 Dataset Collection	29
4.5 Preprocessing	31
4.6 Word Representation	32
4.7 Supervised Learning	33
4.7.1 Logistic Regression	33
4.7.2 SVM	34
4.8 Semi-supervised Learning	35
4.8.1 Self-training	35
4.8.2 Co-Training	37
4.9 Evaluation Design	40
CHAPTER V IMPLEMENTATION	41
5.1 Dataset	44
5.2 Word Representation	45
5.3 Model	47
5.3.1 Self-training	48
5.3.2 Co-training	50
CHAPTER VI RESULTS AND DISCUSSION	54
6.1 Supervised Learning	54
6.2 Semi-supervised Learning	55
CHAPTER VII CONCLUSION	64
7.1 Conclusion	64
7.2 Future Works	65
REFERENCES	66