TABLE OF CONTENTS

APPROVAL PAGE i
SELF DECLARATION AGAINTS PLAGIARISM iii
ABSTRACT iv
ABSTRAK v
DEDICATION & ACKNOWLEDGMENTS vi
PREFACE vii
TABLE OF CONTENTS viii
LIST OF FIGURES x
LIST OF TABLES xi
LIST OF ABBREVIATIONS xii
CHAPTER I INTRODUCTION 1
1.1 Rationale 1
1.2 Statement of Problem and Research Question 3
1.3 Objective and Hypothesis 3
1.3.1 Objective 3
1.3.2 Hypothesis 3
1.4 Assumption 3
1.5 Scope and Delimitation 3
1.5.1 Scope 3
1.5.2 Delimitation 4
1.6 Importance of the Study 4
CHAPTER 2 REVIEW OF LITERATURE AND STUDIES 5
2.1 Related Work 5
2.2 Feature Selection 5
2.2.1 TFIDF 5
2.2.2 Doc2Vec 6
2.2.2.1 Parameters on the Doc2Vec 7
2.2.3 Word2Vec 8
2.3 Classifier Algorithm 8
2.3.1 Rocchio 8
2.3.2 Multinomial Naive Bayes 9
2.3.3 Support Vector Machine 9
2.3.3.1 Separation Linearly 9
2.3.3.2 Soft Margin Hyperplanes 9

2.3.3.3 Kernels 10
2.3.3.4 Parameters in SVM 10
2.3.4 Logistic Regression 11
2.3.5 Gaussian Naive Bayes 12
2.3.6 Bernoulli Naive Bayes 12
2.4 Confusion Matrix 12
CHAPTER 3 RESEARCH METHODOLOGY 14
3.1 The Proposed Approach 14
3.2 Proposed Method 14
3.2.1 Preliminary Evaluation of the Dateset 14
3.2.2 Flow Diagram of the Proposed Method 18
3.3 Measurement 25
CHAPTER 4 DATA EXPERIMENTS AND ANALYSIS 26
4.1 Dataset 26
4.2 Experiment Setting 26
4.2.1 Use Of Preprocessing 26
4.2.2 Parameter in Doc2Vec 28
4.2.3 Classifier Algorithm 29
4.3 Experiment Result and Analysis 29
4.3.1 Baselin Paper Result 29
4.3.2 Use Of Preprocessing 30
4.3.3 Vector Size Parameter in Doc2Vec 32
4.3.4 Window Size Parameter In Doc2Vec 33
4.3.5 Classifier Algorithm 34
4.3.6 C Parameter in SVM 34
4.3.7 Phenomena in the Experiment 35
CHAPTER 5 CONCLUSIONS AND RECOMENDATIONS 36
5.1 Conclusions 36
5.2 Recommendations 36
BIBLIOGRAPHY 37
APPENDIX A 40
Example of Experimental Result 40