

TABLE OF CONTENTS

TABLE OF CONTENTS	iii
LIST OF TABLES	vi
LIST OF FIGURES.....	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 Advantage	4
CHAPTER II LITERATURE REVIEW.....	5
CHAPTER III BASIC THEORY.....	11
3.1 Opinion Mining	11
3.2 Data Preprocessing	11
3.2.1 Punctuation Removal.....	12
3.2.2 Case Folding	12
3.2.3 Contractions Expansion	12
3.2.4 Stemming.....	13
3.2.5 Stop Word Removal.....	14
3.2.6 Tokenization	14
3.3 K-Fold Cross Validation.....	15
3.4 TF-IDF.....	15
3.5 Naive Bayes Algorithm	16
3.6 System Performance Evaluation.....	20

CHAPTER IV SYSTEM ANALYSIS AND DESIGN	23
4.1 System Analysis	23
4.1.1 General System Description	23
4.1.2 System Requirements Analysis	24
4.2 Opinion Mining System Architecture Design	25
4.2.1 Data Acquisition	26
4.2.2 Data Splitting using K-Fold Cross Validation	28
4.2.3 Data Preprocessing	28
4.2.3.1 Punctuation Removal	29
4.2.3.2 Case Folding	30
4.2.3.3 Contractions Expansion	30
4.2.3.4 Tokenization	31
4.2.3.5 Stop Word Removal	31
4.2.3.6 Stemming	32
4.2.4 Term Weighting using TF-IDF	33
4.2.5 Data Classification using Naive Bayes Algorithm	37
4.2.5.1 Training the Naive Bayes Algorithm	37
4.2.5.2 Predicting Classes using Naive Bayes Algorithm	41
4.2.6 System Performance Evaluation	44
4.3 Experiment	44
CHAPTER V IMPLEMENTATION	46
5.1 Hardware and Software Specifications	46
5.2 System Implementation	46
5.2.1 Data Collection	47
5.2.2 Dataset Splitting using K-Fold Cross Validation	49
5.2.3 Data Preprocessing	50

5.2.4 Calculating the TF-IDF Values.....	52
5.2.5 Training the Naive Bayes Algorithm.....	55
5.2.6 Predicting the Classes using Naive Bayes	59
5.2.7 Evaluating System Performance.....	62
CHAPTER VI RESULTS AND DISCUSSION.....	64
6.1 Data Acquisition Results.....	64
6.2 Data Splitting using K-Fold Cross Validation Results.....	65
6.3 Data Preprocessing Results	67
6.4 TF-IDF Weighting Results	69
6.5 Naive Bayes Algorithm Data Classification Results.....	71
6.5.1 Training Stage Results.....	71
6.5.2 Testing Stage Results.....	73
6.5.3 System Performance Evaluation Results	74
CHAPTER VII CONCLUSION AND SUGGESTION	78
7.1 Conclusion.....	78
7.2 Suggestion	78
REFERENCES.....	79
APPENDICES.....	82
APPENDIX 1. Complete Stop Words List from NLTK Library	82
APPENDIX 2. Stop Words List with Stop Word Reduction	82