

TABLE OF CONTENTS

TABLE OF CONTENTS.....	iii
LIST OF TABLES	vi
LIST OF FIGURES	viii
ABSTRACT	x
1 CHAPTER I INTRODUCTION.....	1
1.1 Research Background.....	1
1.2 Research Problem.....	3
1.3 Research Objectives	3
1.4 Research Scope.....	4
1.5 Research Benefits	4
2 CHAPTER II LITERATURE REVIEW.....	5
3 CHAPTER III THEORETICAL FRAMEWORK.....	9
3.1 Sentiment Analysis.....	9
3.1.1 Lexicon based approach.....	10
3.2 Multi-label Classification	11
3.3 Logistic Regression	15
3.4 Customer Satisfaction.....	18
3.5 Twitter	19
3.6 Data Preprocessing	21
3.7 Feature Extraction	22
3.8 Performance Evaluation	23
4 CHAPTER IV RESEARCH METHODOLOGY	27
4.1 Research Description.....	27
4.2 Research Steps.....	29
4.2.1 Dataset Gathering.....	29
4.2.2 Preprocessing Step	30
4.2.3 Labeled data	36

4.2.4	Feature Extraction	37
4.2.5	Separating the Dataset.....	40
4.2.6	Training Step.....	40
4.2.7	Multi Label Classification.....	41
4.2.8	Performance Evaluation	42
5	CHAPTER V IMPLEMENTATION	43
5.1	Dataset Collection	43
5.1.1	Method of Collection	44
5.1.2	Dataset Structure	47
5.2	Pre-Processing	48
5.2.1	Case Folding	51
5.2.2	Non-ASCII character removal	51
5.2.3	Mention Removal, Hashtag Removal and Retweet Removal.....	51
5.2.4	URL Removal, Number Removal and Non-Alphanumeric Removal 52	
5.2.5	Emoticon Removal.....	52
5.2.6	Tokenization.....	53
5.2.7	Normalization.....	54
5.2.8	Negation Handling	55
5.2.9	Stop Word Removal.....	57
5.2.10	Stemming	58
5.3	Labelling.....	59
5.4	Splitting the Dataset	64
5.5	Feature Extraction Implementation	65
5.6	Multi Label Process	66
5.7	Performance Measure	69
6	CHAPTER VI RESULT AND DISCUSSION.....	72
6.1	Data acquisition	72
6.2	Preprocessing step	73
6.3	Manual labelling.....	73
6.4	Splitting of Dataset.....	76
6.5	Feature Extraction Results.....	78

6.6	Multi Label and Performance Measure Result.....	79
6.7	Analysis	89
7	CHAPTER VII CONCLUSION.....	93
7.1	Conclusion.....	93
7.2	Future Works	94
8	BIBLIOGRAPHY	96