

TABLE OF CONTENTS

APPROVAL PAGE	ii
PREFACE	iii
TABLE OF CONTENTS	v
TABLE OF FIGURES	vii
TABLE OF TABLES	viii
TABLE OF EQUATIONS	ix
ABSTRACT	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 Benefit	3
CHAPTER II: LITERATURE REVIEW	4
CHAPTER III: THEORETICAL BASIS	8
3.1 Word Embedding	8
3.1.1 Term Frequency – Inverse Document Frequency (TF-IDF)	8
3.2 Complement Naïve Bayes	9
3.3 Multinomial Naïve Bayes	10
3.4 Preprocessing	11
3.5 K-Fold Cross Validation	11
3.6 Performance Evaluation	13
3.6.1 Accuracy	13
3.6.2 Precision	13
3.6.3 Recall	13
3.6.4 F-Measure	14
CHAPTER IV: RESEARCH METHODOLOGY	15

4.1	Research Description	15
4.2	Research Phases	15
4.3	Research Design	16
4.4	Dataset Collection	17
4.5	Preprocessing	18
4.6	Word Embedding	21
4.7	Complement Naïve Bayes Classifier	22
4.8	Evaluation	24
CHAPTER V: IMPLEMENTATION		26
5.1	System Specification	26
5.2	Dataset	26
5.3	Preprocessing	28
5.2.1	Text Cleansing	28
5.2.2	Lowercasing	29
5.2.3	Punctuation and Digit Removal	29
5.2.4	Stopword Removal	30
5.2.5	Slang Words Conversion	31
5.2.6	Stemming	32
5.4	Word Embedding	32
5.5	Classifiers	33
5.5.1	Complement Naïve Bayes with K-Fold Cross Validation	34
5.5.2	Multinomial Naïve Bayes with K-Fold Cross Validation	35
5.5.3	Complement Naïve Bayes Classification	36
5.5.4	Multinomial Naïve Bayes Classification	36
5.6	Results	37
5.6.1	K-fold Cross Validation	37
5.6.2	CNB and MNB Classifier	38
CHAPTER VI: RESULT AND DISCUSSION		39
6.1	Preprocessing	39
6.2	Preprocessing Result	40



6.3 K-Fold Cross Validation Results	40
6.4 Classification Results	42
CHAPTER VII: CONCLUSION AND SUGGESTIONS	45
7.1 Conclusion	45
7.2 Suggestion	45
REFERENCES	47