

TABLE OF CONTENTS

CHAPTER I: INTRODUCTION	1
1.1 Research Background	1
1.2 Research Problem	4
1.3 Research Scope	5
1.4 Research Objectives	5
1.5 Research Benefits	5
CHAPTER II: LITERATURE REVIEW	6
CHAPTER III: THEORETICAL BASIS	11
3.1 Clickbait	11
3.2 Text Pre-processing	12
3.2.1 Case folding	13
3.2.2 Stopwords removal	13
3.2.3 Punctuations removal	14
3.2.4 Tokenization	15
3.3 Feature Extraction	16
3.3.1 N-grams	16
3.3.2 TF-IDF (Term Frequency - Inverse Document Frequency)	16
3.4 Classification Task	17
3.4.1 Random Forest	17
3.4.2 Naïve Bayes	18
3.5 Performance Evaluation	18
3.5.1 Accuracy	19
3.5.2 Precision	20
3.5.3 Recall	20
3.5.4 F1-Score	20
CHAPTER IV: RESEARCH METHODOLOGY	22

4.1 General Description	22
4.2 Research Phases	23
4.3 Procedure of Data Collection	26
4.3.1 Clickbait Dataset	26
4.3.2 Non-clickbait Dataset	27
4.4 System Design	28
4.4.1 Data Pre-processing	28
4.4.2 Feature Extraction	29
4.4.2 Data Splitting and Algorithm Implementation	29
4.5 Tools and Materials	32
4.5.1 Tools	32
4.5.2 Materials	32
CHAPTER V: IMPLEMENTATION	33
5.1 Implementation of Data Collection and Labelling	33
5.1.1 Clickbait Dataset	35
5.1.2 Non-clickbait Dataset	35
5.2 Implementation of Data Pre-processing	36
5.3 Implementation of Feature Extraction	38
5.3.1 TFIDF and Unigram	39
5.3.2 TFIDF and Bigram	39
5.4 Implementation of Train and Test Dataset Splitting	40
5.5 Implementation of Random Forest	41
5.5.1 Random Forest without Hyperparameter	42
5.5.2 Random Forest with Hyperparameter	42
5.6 Implementation of Evaluation	46
5.6.1 Implementation of Classification Report Calculation	46
5.6.2 Implementation of Cross Validation	47
5.6.3 Implementation of Naïve Bayes	48

CHAPTER VI: RESULTS AND DISCUSSIONS	50
6.1 Data Collection and Labelling Results	50
6.1.1 Clickbait Dataset Results	50
6.1.2 Non-clickbait Dataset Results	51
6.2 Data Pre-processing Results	51
6.3 Feature Extraction Results	52
6.3.1 TFIDF and Unigram Results	53
6.3.2 TFIDF and Bigram Results	54
6.4 Dataset Splitting Results	55
6.5 Classification Results	57
6.6 Performance Evaluation Results	60
6.6.1 Classification Report	60
6.6.2 Cross Validation Results	63
6.6.3 Comparison with Naïve Bayes Results	64
6.6.3 Performance Evaluation Results with Another Dataset	65
CHAPTER VII: CONCLUSIONS AND RECOMMENDATIONS	68
7.1 Conclusions	68
7.2 Recommendations	68
REFERENCES	70