

## TABLE OF CONTENTS

<b>APPROVAL PAGE</b>	<b>iii</b>
<b>STATEMENT</b>	<b>iv</b>
<b>FOREWORD</b>	<b>vi</b>
<b>TABLE OF CONTENTS</b>	<b>vii</b>
<b>LIST OF FIGURES</b>	<b>xi</b>
<b>LIST OF TABLES</b>	<b>xiii</b>
<b>LIST OF LISTINGS</b>	<b>xiv</b>
<b>ABSTRACT</b>	<b>xv</b>
<b>CHAPTER I INTRODUCTION</b>	<b>1</b>
<b>1.1 Research Background</b>	<b>1</b>
<b>1.2 Research Problem</b>	<b>5</b>
<b>1.3 Research Scope</b>	<b>5</b>
<b>1.4 Research Objective</b>	<b>5</b>
<b>1.5 Research Benefits</b>	<b>6</b>
<b>1.6 Research Methodology</b>	<b>6</b>
<b>1.7 Thesis Outline</b>	<b>8</b>
<b>CHAPTER II LITERATURE REVIEW</b>	<b>10</b>
<b>CHAPTER III THEORETICAL BASIS</b>	<b>16</b>
<b>3.1 Text Representation</b>	<b>16</b>
3.1.1 Word2vec	19
3.1.2 FastText	20
<b>3.2 RNN</b>	<b>20</b>
3.2.1 Forward Pass	21



3.2.2 Backward Pass	22
3.2.3 Bidirectional RNN	23
3.2.4 Limitations of RNN	23
<b>3.3 LSTM</b>	<b>24</b>
3.3.1 Bidirectional LSTM	26
<b>3.4 CNN</b>	<b>27</b>
3.4.1 CNN for Text Classification	27
<b>3.5 Performance Evaluation</b>	<b>28</b>
3.5.1 Accuracy	28
3.5.2 Precision	29
3.5.3 Recall	29
3.5.4 F1 Score	29
<b>3.6 Fleiss' Kappa</b>	<b>30</b>
<b>CHAPTER IV ANALYSIS AND SYSTEM DESIGN</b>	<b>31</b>
<b>4.1 Research Description</b>	<b>31</b>
<b>4.2 Research Tools</b>	<b>31</b>
<b>4.3 Research Phases</b>	<b>33</b>
<b>4.4 Dataset Collection</b>	<b>37</b>
<b>4.5 Data Pre-processing</b>	<b>38</b>
<b>4.6 Model Design</b>	<b>39</b>
4.6.1 Bi-LSTM	39
4.6.2 CNN	40
<b>4.7 Evaluation and Analysis Design</b>	<b>41</b>
<b>CHAPTER V IMPLEMENTATION</b>	<b>43</b>
<b>5.1 Dataset Collection</b>	<b>43</b>

5.1.1 Method of Collection	45
5.1.2 Dataset Structure	47
<b>5.2 Labelling</b>	<b>48</b>
<b>5.3 Dataset Pre-processing</b>	<b>52</b>
<b>5.4 Dataset Operation</b>	<b>53</b>
5.4.1 Training, Validation, Test Data	53
5.4.2 Dataset Visualization	54
5.4.3 Dataset pre-process	55
<b>5.5 Model Implementation</b>	<b>59</b>
5.5.1 Evaluation Metrics	59
5.5.2 Embedding Layer	59
5.5.3 Model Bi-LSTM	61
5.5.4 Model CNN	64
5.5.5 K-fold Validation	69
5.5.6 Testing	70
5.5.7 Lime Explainer Implementation	71
<b>CHAPTER VI RESULTS AND DISCUSSION</b>	<b>73</b>
<b>6.1 CLICK-ID: Clickbait Dataset</b>	<b>73</b>
6.1.1 Scraping Result	73
6.1.2 Analysis of inter-annotator agreement on dataset	75
<b>6.2 Analysis per Publisher</b>	<b>79</b>
<b>6.3 Category Comparison</b>	<b>82</b>
6.3.1 Clickbait Comparison for News Category Group	84
6.3.2 Clickbait Comparison for Celebrity&Entertainment Category	86
6.3.3 Other Category Group Comparison	87



<b>6.4 Dataset Length Analysis</b>	<b>91</b>
<b>6.5 Dataset Division</b>	<b>92</b>
<b>6.6 K-Fold Validation</b>	<b>94</b>
<b>6.7 Model Performance</b>	<b>96</b>
<b>6.8 LIME Explainer</b>	<b>98</b>
<b>6.9 Further Discussion</b>	<b>101</b>
6.9.1 Why only headlines were consider not the content?	101
6.9.2 Does clickbait headlines in newspapers still considered as a 'clickbait'?	101
<b>CHAPTER VII CONCLUSION</b>	<b>103</b>
<b>7.1 Conclusion</b>	<b>103</b>
<b>7.2 Future work</b>	<b>104</b>
<b>REFERENCES</b>	<b>106</b>