

CONTENT

ABSTRACT.....	2
CONTENT.....	3
CHAPTER 1: INTRODUCTION.....	1
1.1. Introduction.....	1
Table 1.1 List of Dental Specialties recognized by ADA.....	1
1.2. Problem Definition.....	2
1.3. Objective & Hypothesis.....	3
1.4. Assumption.....	3
1.5. Scope and Limitation.....	4
CHAPTER 2: STUDY LITERATURE.....	5
2.1. Dental Chief Complaint.....	5
2.2. About Text Classification.....	5
Figure 2.1 Natural Language Processing (NLP) Tasks.....	5
2.3. Complaint Text Pre-processing.....	6
Table 2.1 Steps taken in text pre-processing.....	6
Figure 2.2 Stemming vs Lemmatization.....	7
2.4. Word Embedding.....	8
Figure 2.3 Man - Woman vector.....	8
Figure 2.4 Comparative - Superlative vector.....	9
Figure 2.5 Keras Tokenizer.....	9
Figure 2.6 FastText model illustration.....	10
Figure 2.7 BERT input representation.....	10
2.5. Observation detail about CNN, Bi-LSTM and Attention layer.....	12
Figure 2.8 Illustration of CNN Architecture.....	12
Figure 2.9 Illustration of Bi-LSTM Architecture.....	13
Figure 2.10 Illustration of Attention diagram taken from Wikipedia.....	14
2.6. Proposed Classification Method.....	15
Figure 2.11 The Bi-LSTM architecture with its visualization of attention mechanism.....	15
Figure 2.12 Architecture Word2Vec combine with convolution and max polling process.....	16
Figure 2.13 Complete architecture of combination Word2Vec, CNN and Bi-LSTM.....	16
Figure 2.14 the CNN model structure.....	17
CHAPTER 3: RESEARCH AND METHODOLOGY.....	18
3.1. Research Design.....	18
Figure 3.1 Research Workflow.....	18
Table 3.1 Sample of Raw Data with its label.....	19
Figure 3.2 Flowchart of Sentence Filtering Process.....	20
Table 3.2 Example Normal Dataset and Sentence Filtering dataset.....	21
Figure 3.3 Formula for Precision and Recall.....	22
Figure 3.4 Formula for F1 Score.....	22
3.2. Model Architecture.....	23

Figure 3.5 CNN architecture.....	23
Figure 3.6 Bi-LSTM architecture.....	23
Figure 3.7 Multinomial Naive Bayes architecture.....	24
3.3. Experiment Scenario.....	24
Table 3.3 List of experiment models and its configuration.....	24
CHAPTER 4: EXPERIMENT AND ANALYSIS.....	25
4.1. Dataset.....	25
Table 4.1 Dataset distribution for each label.....	25
Figure 4.1 Code for Train-Test splitting.....	25
4.2. Experimental Configuration.....	26
4.3. Experimental Results.....	26
Figure 4.2 Model Comparison that use Sentence Filtering dataset.....	26
Figure 4.3 Model Comparison that use Normal dataset.....	27
Figure 4.4 Confusion matrix of model CNN.....	28
Figure 4.5 The word-cloud of Oral Pathology, Orthodontics and Prosthodontics.....	28
CHAPTER 5: CONCLUSION AND RECOMMENDATION.....	29
5.1. Conclusion.....	29
Figure 5.1 Sample dataset from previous journal experiment.....	30
Table 5.1 The word-cloud characteristics for 6 (six) dental category.....	30
5.2. Recommendation.....	31
BIBLIOGRAPHY.....	32