Table of Contents

Abstract i
Table of Contents ii
1.0 Introduction 1
Background 3
Natural Language Processing 3
Text Extraction 3
Textrank 3
TF-IDF 4
Text Abstraction 4
Artificial Neural Network 5
RNN and LSTM 5
Word Embedding 8
ROUGE-N and BLEU Metrics 8
Libraries 9
Keras 9
NLTK 9
Scikit-learn 9
Pandas 10
Gensim 10
Flask 10
Bootstrap 10
GloVe 10
LXML 11
3.0 Related Work 12
Methodology 15
Choose and clean datasets 15
Datasets Information 15
Data Cleaning 17
Data Categorization 19
Extractive Summarization 21

Algorithms 21
Control Experiment 22
Metrics 22
Abstractive Summarization 22
Preparing the dataset 24
Embedding layer 25
Model 26
Test and compare among different datasets 29
Tune the model 29
Build an end-to-end application 30
Results 31
Data Cleaning and Categorizing 31
Extractive Model Performance 33
Abstractive Model Performance 35
Performance Comparison Among Datasets 46
End-to-End Application 47
6.0 Conclusion 49
Reference 51