UNIVERSITAS GADJAH MADA
FINE-GRAINED SENTIMENT CLASSIFICATION ON COVID-19 VACCINATION IN INDONESIA USING BIDIRECTIONAL
ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT) AND CONVOLUTIONAL NEURAL NETWORK (CNN)
FEDORA R W P, Azhari, Drs., M.T., Dr
Universitas Gadjah Mada, 2022 | Diunduh dari http://etd.repository.ugm.ac.id/

TABLE OF CONTENTS

APPROVAL PAGE iii
STATEMENT PAGE iv
FOREWORD v
TABLE OF CONTENTS vii
TABLE OF TABLES x
TABLE OF FIGURES xi
ABSTRACT xiii
ABSTRAK xiv
CHAPTER I 1
INTRODUCTION 1
1.1 Research Background 1
1.2 Research Problem 3
1.3 Research Objective 3
1.4 Research Scope 4
1.5 Research Benefit 4
CHAPTER II 5
LITERATURE REVIEW 5
CHAPTER III 9
FUNDAMENTAL THEORY 9
3.1 Text Mining 9
3.2 Fine-grained Sentiment Classification 10
3.3 Preprocessing 10
3.3.1 Case Folding 11
3.3.2 Stop Word Removal 11
3.3.3 Special Word Removal 11
3.3.4 Stemming 12
3.3.5 Normalization 12
3.4 Feature Extraction 13
3.4.1 Word Embedding 13
3.4.2 Word2Vec 13
3.4.3 Contextualized Word Embedding 15

UNIVERSITAS GADJAH MADA
FINE-GRAINED SENTIMENT CLASSIFICATION ON COVID-19 VACCINATION IN INDONESIA USING BIDIRECTIONAL
ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT) AND CONVOLUTIONAL NEURAL NETWORK (CNN)
FEDORA R W P, Azhari, Drs., M.T., Dr
Universitas Gadjah Mada, 2022 | Diunduh dari http://etd.repository.ugm.ac.id/

3.4.4 BERT 16
3.4.5 IndoBERT 17
3.5 Convolutional Neural Network 17
3.5.1 Convolutional Layer 18
3.5.2 Pooling Layer 18
3.5.3 Fully Connected (FC) Layer 19
3.6 Model Evaluation 19
CHAPTER IV 22
RESEARCH METHODOLOGY 22
4.1 Research Description 22
4.2 Tools and Materials 23
4.3 Data Collection 23
4.4 Data Labelling 23
4.5 Data Splitting 24
4.6 Preprocessing 24
4.7 Feature Extraction 25
4.8 Model Training and Hyperparameter Tuning 26
4.9 Model Evaluation 28
CHAPTER V 29
IMPLEMENTATION 29
5.1 Data Labelling 29
5.2 Preprocessing Implementation 29
5.2.1 Casefolding 29
5.2.2 Special Word Removal 30
5.2.3 Stopword Removal 30
5.2.4 Normalization 30
5.2.5 Stemming 31
5.3 Split Data 31
5.4 Feature Extraction Implementation 31
5.4.1 Word2Vec 31
5.4.2 BERT 31
5.5 Sentiment Classification Implementation 32
5.5.1 Data Embedding 32

UNIVERSITAS
GADJAH MADA
FINE-GRAINED SENTIMENT CLASSIFICATION ON COVID-19 VACCINATION IN INDONESIA USING BIDIRECTIONAL
ENCODER REPRESENTATIONS FROM TRANSFORMERS (BERT) AND CONVOLUTIONAL NEURAL NETWORK (CNN)
FEDORA R W P, Azhari, Drs., M.T., Dr
Universitas Gadjah Mada, 2022 | Diunduh dari http://etd.repository.ugm.ac.id/

5.5.2 CNN Model 34
5.6 Hyperparameter Tuning 36
5.7 Model Evaluation 37
CHAPTER VI 38
RESULT AND DISCUSSION 38
6.1 Labeling Data Result 38
6.2 Splitting Data Result 38
6.3 Hyperparameter Tuning Result 40
6.4 Validation Result 40
6.5 CNN Model Test Results 41
CHAPTER VII 45
CONCLUSIONS AND SUGGESTIONS 45
7.1 Conclusions 45
7.2 Suggestions 45
REFERENCES 46