
CONTENTS

APPROVAL.....	i
SELF DECLARATION AGAINTS PLAGIARISM.....	ii
ABSTRACT	iii
ABSTRAK	iv
DEDICATION	v
ACKNOWLEDGEMENTS	vi
CONTENTS	vii
LIST OF FIGURES	xi
LIST OF TABLES.....	xii
CHAPTER 1 INTRODUCTION.....	1
1. 1 Rationale	1
1. 2 Theoretical Framework.....	3
1. 3 Conceptual Framework.....	4
1. 4 Problems Statements.....	4
1. 5 Objective	5
1. 6 Hypotheses	5
1. 7 Scope and Delimitation.....	6
1. 8 Importance of the Study.....	7
CHAPTER 2 REVIEW OF LITERATURE AND STUDIES	8
2. 1 Personality Prediction using Machine Learning	8
2. 2 The Big Five Personality Model	12
2. 3 User Information and Grammatical Features in Twitter.....	13
2. 4 Personality Prediction Methodology	16
2. 4. 1 Stacking.....	16

2. 4. 2	Random Forest	17
2. 4. 3	Gradient Boosting.....	18
2. 4. 4	K-Nearest Neighbor (K-NN).....	19
2. 5	Machine Learning Optimization Technique	20
2. 5. 1	Chi-Square for Feature Selection Technique	20
2. 5. 2	Synthetic Minority Over-Sampling (SMOTE) for Sampling Technique	21
2. 5. 3	Grid Search for Hyper-parameter Optimization Technique	23
2. 6	Data Representation.....	24
2. 6. 1	N-gram	24
2. 6. 2	Term Frequency-Inverse Document Frequency (TF-IDF)	24
2. 7	K-fold Cross Validation.....	25
2. 8	Performance Evaluation.....	25
CHAPTER 3 RESEARCH METHODOLOGY		27
3. 1	Research Design	28
3. 1. 1	Data Preprocessing	29
3. 1. 2	Data Representation.....	32
3. 1. 3	Feature Selection	33
3. 1. 4	Sampling	34
3. 1. 5	Hyper-parameters Optimization	35
3. 1. 6	Processing	36
3. 1. 7	Post-Processing (Performance Evaluation).....	41
3. 2	System Implementation	41
3.2.1	Hardware Specifications	41
3.2.2	Software Specifications	41

3.3	Dataset and Features	41
3.4	Experiment Design	45
CHAPTER 4 RESULT AND ANALYSIS		48
4.1	Presentation of Data.....	48
4.2	Experiment Result	49
4.2.1	The Result of Personality Prediction using K-NN, Random Forest and Gradient Boosting (1 st Scenario).....	50
4.2.2	The Result of Personality Prediction using Stacking (2 nd Scenario).....	52
4.2.3	The Result of Personality Prediction using Modified Stacking (3 rd Scenario).....	53
4.2.4	The Result of Personality Prediction using Stacking and Modified Stacking without Grammatical Features (4 th Scenario).....	54
4.3	Performance Comparison.....	55
4.3.1	1 st Comparison Performance of K-NN, Random Forest, Gradient Boosting, Stacking and Modified Stacking method	55
4.3.2	2 nd Comparison Performance of Stacking and Modified Stacking Method using All Feature with Stacking and Modified Stacking Method without Grammatical Features.....	59
4.3.3	3 rd Comparison Performance of Modified Stacking Method with Similar Research.	60
CHAPTER 5 CONCLUSION AND RECOMMENDATIONS.....		62
5.1	Conclusion.....	62
5.2	Recommendations	63
REFERENCES		64
APPENDIX A List of Emoticons.....		67
APPENDIX B Negative and Positive Words in Bahasa.....		74

APPENDIX C Evaluation Result of Unigram and Bigram 92

APPENDIX D Evaluation Result of Classifier in Feature Selection 93

APPENDIX E Hyper-paramater Optimization Result..... 94