

CONTENTS

APPROVAL PAGE..... ii

SELF DECLARATION AGAINST PLAGIARISM iii

ABSTRACT iv

ABSTRAK v

DEDICATION..... vi

ACKNOWLEDGMENTS..... vii

CONTENTS.....viii

LIST OF TABLES xi

LIST OF FIGURES.....xii

CHAPTER 1: THE PROBLEM..... 1

 1.1 Background 1

 1.2 Problem Identification 2

 1.3 Scope of Work 3

 1.4 Objective..... 3

 1.5 Hypothesis 3

CHAPTER 2: REVIEW OF LITERATURE AND STUDIES..... 4

 2.1 Previous Studies..... 4

 2.2 Preprocessing 6

 2.2.1 Missing Value..... 7

 2.2.2 Imbalanced Data..... 7

 2.2.3 Rank Feature..... 7

 2.2.4 Data Augmentation 7

CHAPTER 3: RESEARCH METHODOLOGY 8

 3.1 Main Contributions..... 8

 3.2 Research Design..... 9

 3.2.1 Data Collecting Process 9

 3.2.2 Data Preprocessing..... 9

 3.2.2.1 Missing Value..... 9

 3.2.2.2 Imbalanced Data..... 10

 3.2.2.3 Feature Selection (Rank Features)..... 10

3.2.2.4	Data Augmentation	10
3.2.3	Classification	10
3.2.4	Model Validation	11
3.2.5	Result Analysis	11
3.3	Experiment Design.....	11
CHAPTER 4: ANALYSIS AND INTERPRETATION OF DATA		12
4.1	Dataset.....	12
4.1.1	Pima Indian Dataset.....	12
4.1.2	Karya Medika Dataset.....	13
4.2	Data Exploration	14
4.2.1	Pima Indian Dataset.....	15
4.2.2	Karya Medika Dataset.....	17
4.2.3	Baseline difference in 2 Dataset	19
4.3	Data Preprocessing.....	22
4.3.1	Missing Value.....	22
4.3.2	Imbalanced Data.....	24
4.3.3	Feature Selection (Rank Features).....	27
4.3.4	Data Augmentation	30
4.4	Classification	32
4.4.1	Random Forest.....	32
4.4.2	Logistic Regression.....	34
CHAPTER 5: EXPERIMENT AND ANALYSIS RESULT		36
5.1	Experiment Result.....	36
5.1.1	Missing Value.....	36
5.1.2	Imbalanced Data.....	36
5.1.3	Feature Importance (Rank Feature)	37
5.1.4	Classification	38
5.2	Analysis Result	38
5.2.1	Missing Value.....	39
5.2.2	Imbalanced Data.....	40
5.2.3	Feature Importance (Rank Feature)	41
5.2.4	Summary of Experiments	41
5.2.5	Classification Result	42

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS.....	46
6.1 Conclusions.....	46
6.2 Recommendations.....	46
6.3 Future Work.....	46
BIBLIOGRAPHY	47
CURRICULUM VITAE	50