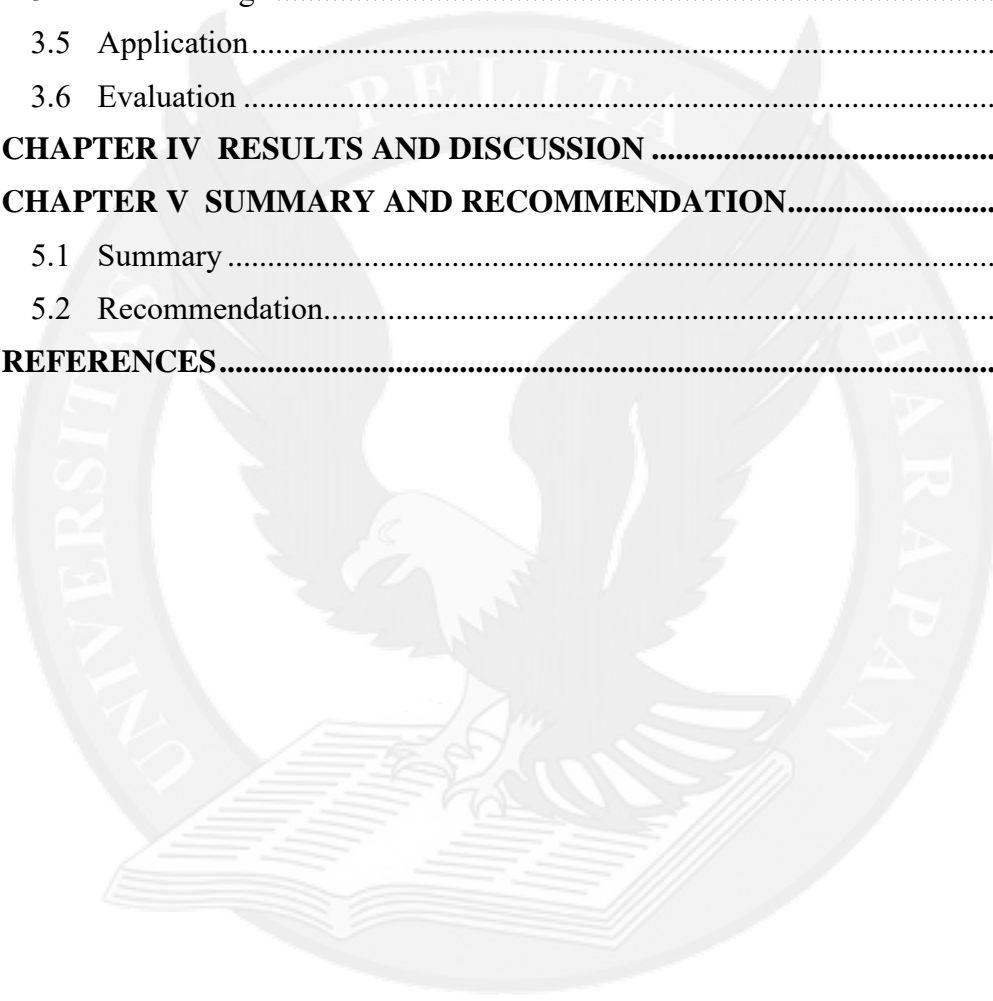


TABLE OF CONTENTS

COVER	
ORIGINALITY STATEMENT OF THE THESIS	
THESIS SUPERVISOR'S APPROVAL	
THESIS EXAMINATION COMMITTEE	
ABSTRACT	v
ACKNOWLEDGEMENTS	vii
TABLE OF CONTENTS	ix
LIST OF FIGURES	xi
LIST OF TABLES	xii
CHAPTER I INTRODUCTION	1
1.1 Background	1
1.2 Problem <i>Identification</i>	6
1.3 Problem <i>Limitation</i>	7
1.4 Problem Formulation	7
1.5 Research Objectives	8
1.6 Thesis Outline	8
CHAPTER II LITERATURE REVIEW	10
2.1 General Understanding of Loan Default.	10
2.2 Data <i>Standardization</i>	11
2.3 Machine Learning Theory.....	12
2.3.1 Random Forest.....	16
2.3.2 Decision Tree.....	17
2.3.3 Gradient Boosting.....	19
2.3.4 Naïve Bayes.....	20
2.4 Testing <i>and Assessment</i>	21
2.4.1 Testing	22
2.4.1.1 Training Data.....	22
2.4.1.2 Test Data	23
2.4.2 Evaluation.....	24
2.4.2.1 Confusion Matrix.....	24
2.4.2.2 Accuracy	26
2.4.2.3 Precision	27

2.4.2.4 Recall.....	27
2.4.2.5 Receiver Operating Characteristic (ROC).....	28
2.4.2.6 Area Under Curve (AUC)	29
CHAPTER III RESEARCH METHODOLOGY	31
3.1 Overview	31
3.2 Data Preparation.....	33
3.3 Pre-processing.....	35
3.4 Model Usage	37
3.5 Application.....	38
3.6 Evaluation	39
CHAPTER IV RESULTS AND DISCUSSION	41
CHAPTER V SUMMARY AND RECOMMENDATION.....	50
5.1 Summary	50
5.2 Recommendation.....	51
REFERENCES.....	53



LIST OF FIGURES

Figure 2. 1 ROC Graph	29
Figure 3. 1 Research Block Diagram	31
Figure 3. 2 Pre-Processing Table	35



LIST OF TABLES

Table 2. 1 Confusion Matrix	25
Table 4. 1 ADDR attribute sensitivity results	48
Table 4. 2 UMUR attribute sensitivity results	48
Table 4. 3 SEX_DESC attribute sensitivity results.....	48
Table 4. 4 KD_PRODUK attribute sensitivity results	49
Table 4. 5 NM_PROV attribute sensitivity results	49

