



TABLE OF CONTENTS

APPROVAL PAGE	iii
STATEMENT	iv
DEDICATION PAGE	v
QUOTATION PAGE	vi
FOREWORD	vii
TABLE OF CONTENTS	ix
LIST OF FIGURES	xii
LIST OF TABLES	xiii
ABSTRACT	xv
INTISARI	xvi
CHAPTER I: INTRODUCTION	1
1.1 Background.....	1
1.2 Research Question	2
1.3 Research Scope.....	2
1.4 Research Objective.....	3
1.5 Thesis Organization.....	3
CHAPTER II: LITERATURE REVIEW	5
CHAPTER III: THEORITICAL BASIS	12
3.1 Machine Learning.....	12
3.1.1 Machine Learning Definition.....	12
3.1.2 Machine Learning Process.....	12
3.1.3 Types of Machine Learning.....	13
3.2. Modeling Design.....	14
3.2.1 Problem Formulation.....	14
3.2.2 Data Cleansing.....	14
3.2.3 Data Analysis.....	15
3.2.4 Feature Processing.....	16
3.2.5 Dimensionality Reduction.....	16
3.2.6 Data Training.....	17
3.2.7 Data Testing.....	17



3.3	Diabetes Mellitus Type 2.....	18
3.4	ICD-9 Code.....	19
3.5	Guid Data.....	20
3.6	Naive Bayes.....	20
3.6.1	Bayesian Decision Theory.....	20
3.6.2	Naive Bayes Definition.....	21
3.6.3	Algorithm.....	21
3.6.4	Advantages and Disadvantages.....	24
3.7	K Nearest Neighbor.....	25
3.7.1	K Nearest Neighbor Definition.....	25
3.7.2	Algorithm.....	25
3.7.3	Advantages and Disadvantages.....	26
3.8	Random Forest.....	26
3.8.1	Random Forest Definition.....	26
3.8.2	Algorithm.....	27
3.8.3	Advantages and Disadvantages.....	28
3.9	Ensemble Learning Weighted Voting.....	29
3.10	Performance Calculations.....	29
3.10.1	Precision and Recall.....	29
3.10.2	Accuracy.....	30
	CHAPTER IV: SYSTEM DESIGN.....	31
4.1	Research Process.....	31
4.2	Data Collection.....	32
4.2.1	Data Cleansing.....	38
4.2.2	Data Analysis.....	39
4.3	Feature Processing.....	40
4.4	Dimensionality Reduction.....	41
4.4.1	Feature Selection.....	41
4.4.2	Feature Extraction.....	42
4.3.3	Normalization.....	43
4.5	Data Training.....	44
4.6	Data Testing.....	51
4.7	Systems Interface Design.....	52



CHAPTER V: IMPLEMENTATION	54
5.1 Data Cleansing.....	54
5.2 Data Analysis.....	55
5.3 Feature Processing.....	56
5.4 Dimensionality Reduction.....	59
5.4.1 Feature Selection.....	59
5.4.2 Feature Extraction.....	60
5.4.3 Normalization.....	61
5.5 Data Training.....	62
5.6 Data Testing.....	64
5.6.1 Naive Bayes.....	64
5.6.2 K Nearest Neighbor.....	64
5.6.3 Random Forest.....	65
5.6.4 Ensemble Learning Weighted Voting.....	65
5.7 Systems Interface Implementation.....	68
CHAPTER VI: EXPERIMENT AND DISCUSSION	74
6.1 Experiment for Naive Bayes, K Nearest Neighbor, Random Forest.....	74
6.2 Random Forest and Ensemble Learning Weighted Voting Parameters for Precision, Recall and Accuracy.....	75
6.3. Algorithm Comparison with Ensemble Learning in Precision, Recall and Accuracy.....	81
CHAPTER VII: CONCLUSION	82
7.1 Conclusion.....	82
7.2 Future Work.....	82
BIBLIOGRAPHY	83
APPENDICES	87