

TABLE OF CONTENTS

CHAPTER NO.	TITLE	PAGE NO.
	ABSTRACT.....	v
	TABLE OF CONTENTS.....	vii
	LIST OF FIGURES.....	xiii
	LIST OF TABLES.....	xvi
	LIST OF SYMBOLS AND ABBREVIATIONS.....	xvii
1	INTRODUCTION.....	1
1.1	Overview.	1
1.2	Data Mining and Knowledge Discovery.....	6
1.3	Applications of Data Mining.....	8
1.4	Challenges in Data Mining.....	9
1.5	Analysis of the Data Mining.....	11
1.6	Scientific Contributions.....	17
1.7	Kind of Information to be Collected.....	18
1.8	Kind of Data to be Mined.....	20
1.9	Classification and Regression.....	26
1.10	Bias and Variance Decomposition.....	27
	1.10.1 Measuring Bias and Variance.....	28
1.11	Ensemble Learning.....	28
	1.11.1 Ensemble Learning Algorithms.....	29
1.12	Problem Domains.....	31
	1.12.1 Intrusion Detection System.....	31
	1.12.2 Direct Marketing.....	32
	1.12.3 Signature Verification.....	34
	1.12.4 Overview	38
	1.12.5 Classification of Data Mining	39
	1.12.6 Architecture of Data Mining	41
	1.12.7 Data Mining Life cycle.....	43
1.13	Motivation.....	44
1.14	Problem Specification.....	46
1.15	Dissertation Objectives	46

1.16	Scope and Objective of Thesis	47
1.17	Organization of the Thesis	48
2	LITERATURE REVIEW.....	49
2.1	Evolution of Data Mining Techniques.....	49
2.2	Association Rule Mining.....	49
2.2.1	Techniques based on Horizontal Layout of Databases	50
2.2.2	Techniques in view of Vertical Layout of Databases	52
2.2.3	Techniques in view of Projected Layout of Database.....	53
2.2.4	Graph-Based Approaches of Rule Mining.....	54
2.2.5	Advanced Approaches of Rule Mining.....	55
2.2.6	Techniques that Enhance the Efficiency of Rule Mining	56
2.3	Swarm Intelligence.....	58
2.3.1	Ant Colony Optimization	58
2.3.2	Particle Swarm Optimization (PSO)	59
2.3.3	Other Swarm Intelligence Techniques	60
2.4	Data Mining with Ant Colony Optimization	61
2.4.1	Classification Rule Discovery with Ant Miner	61
2.4.2	Clustering with ACO.....	63
2.4.3	Association Rule mining with ACO.....	64
2.5	Background Studies.....	64
2.6	Inference from Existing Work.....	64
2.7	Recent Studies.....	65
2.8	Summary	74
3	ASSOCIATION RULE MINING AND REGRESSION.....	76
3.1	Overview of Association Rule Mining (ARM.).....	76
3.1.1	Types of Association Rules.....	79
3.1.2	Increasing the Efficiency of Association Rules Algorithm	81
3.2	Apriori Algorithm.....	81
3.2.1	Apriori Rule Mining	82
3.2.2	Apriori Properties	82
3.2.3	Apriori Approach.....	83
3.2.4	Common Steps in Apriori Algorithm.....	84

3.2.5	Formal Problem Description	87
3.2.6	Frequent Itemset Generation	88
3.2.7	Rule Generation	89
3.2.8	Traversing the Search Space	90
3.2.9	Defining Support and Confidence	91
3.3	Algorithms for Association Rule Mining.....	92
3.3.1	A priori Algorithm.....	93
3.3.2	DHP Algorithm	97
3.3.3	DIC Algorithm	97
3.3.4	Partitioning Algorithm	98
3.3.5	Sampling Algorithm.....	99
3.3.6	Equivalence Class Transformation (ECLAT) Algorithm	100
3.3.7	Log ECLAT Algorithm.....	100
3.3.8	FP-tree/FP-Growth	101
3.3.9	Co-Occurrence Frequent Item-Tree (COFI-Tree) Algorithm	103
3.3.10	CT-PRO Algorithm	104
3.3.11	H-Mine Algorithm.....	104
3.3.12	Rapid Association Rule Mining (RARM) Algorithm.....	105
3.4	Increasing the Efficiency of ARM.....	106
3.5	Recent Advances in Association Rule Discovery.....	107
3.5.1	Redundant Association Rules.....	107
3.5.2	Negative Association Rules	107
3.5.3	Alternate measures of interestingness	108
3.6	Proposed Graph Based Approach for Finding the Frequent Item sets to Discover Association Rules	108
3.6.1	Overview	108
3.6.2	Proposed Approach	109
3.6.3	Graph Construction Phase	109
3.6.4	Generation of Item sets	110
3.6.5	Finding the Frequent Itemsets	111
3.6.6	Efficiency of the Proposed Approach and Comparison with Existing Approach	112
3.7	Applications of Association Rule Mining	112
3.7.1	Market Basket Analysis	112

3.7.2	Medical Diagnosis	114
3.7.3	Census Data	115
3.7.4	Agriculture	115
3.7.5	Other Application Areas	115
3.8	Regression	116
3.8.1	Regression Models	116
3.8.2	Regression Workflow	118
3.8.3	Types of Regression	118
3.8.4	Difference Between Linear and Non-Linear Regression	121
3.8.5	Some other types of Regression	122
3.8.6	Common Applications of Regression	123
3.9	Proposed Methodology	123
3.9.1	Problem Specification	123
3.9.2	Proposed Methodology	124
3.10	Summary	126
4	CLASSIFICATION OF SINGLE VALUED CLASS ATTRIBUTES IN ASSOCIATION RULES MINING.....	128
4.1	Classification based on Association Rules (CBA).....	128
4.2	Post Classification of Association Rules.....	130
4.3	Association Rule Based on Classification Model	131
4.3.1	Classification Association Rules Generation.....	131
4.3.2	Generating Rules with Semantic Constraint.....	132
4.3.3	Classification Models	138
4.3.4	Single Rule and Multiple Rules Classification.....	139
4.4	Implementation	140
4.5	Experimental Evaluation.....	144
4.5.1	Evaluation Metrics.....	144
4.5.2	Experimental Results.....	146
4.6	Adaptive Minimum Support	150
4.7	Initial Min Support Selection	151
4.8	Adaptive Minimal Support Algorithm	152
4.9	Experiments	154
4.9.1	Experiment Design	154

5	CLASSIFICATION OF MULTI-VALUED ATTRIBUTES THROUGH ASSOCIATION RULE MINING	156
5.1	Association Rule Mining by Set-Valued Attributes.....	156
5.2	Classification by Set-Valued Class Attribute.....	156
5.2.1	Set-Valued Class Prediction	156
5.2.2	E-Measure	157
5.3	Building Classification Models	158
5.3.1	SCBA Algorithm	158
5.3.2	All Rules Model	160
5.3.3	Model Prediction	160
5.4	Experimental Evaluation.....	161
6	CONCLUSION	166
	REFERENCES	168
	LIST OF PUBLICATIONS	180
	PLAGIARISM REPORT	187
	VITAE	188