

TABLE OF CONTENTS

CHAPTER NO	PAGE NO
CHAPTER 1. INTRODUCTION	
1.1 Introduction to Machine Learning	1
1.2 Importance of Classification	2
1.3 Motivation	2
1.4 Problem Statement	3
1.4.1 Research Gaps	3
1.4.2 Solution	4
1.5 Research Objectives	4
1.6 Data Description	5
1.7 Organization of Thesis	6
1.8 Summary Of The Proposed Framework	7
CHAPTER 2. LITERATURE SURVEY	
2.1 Introduction	8
2.2 Anomaly Detection Models in Different Applications	8
2.3 Traditional Statistical Ranking Models for Classification Problem	10
2.4 Issues of Traditional Statistical Models on Medical Databases	11
2.5 Traditional Feature Selection Based Classification Problems on Medical and KDD Databases	12
2.6 Ensemble Feature Selection Models On Credit Card Databases	13
2.7 Traditional Decision Tree Models On High Dimensional Databases	13
2.8 Traditional Heterogeneous Statistical Data Anomaly Prediction Models	15
2.9 Traditional Pattern Mining Approaches	17
2.10 Traditional Meta Heuristic Feature Selection Based Classification Models	19
2.11 Boosting Classification Models for High Dimensional Databases	23

2.12	Partitioning Based Classification Models for Real-Time Databases	26
2.13	High Dimensional Distance Based Classification Models	29
CHAPTER 3. PROBABILISTIC FEATURE ENGINEERING		
3.1	Introduction	32
3.2	Statistical Feature Extraction in High Dimensional Data	32
3.3	Statistical Control Chart Model	34
	3.3.1 Statistical Control Chart Algorithm's Basic Flow Structure	34
	3.3.2 Modified Decision Tree Model	35
3.4	Experimental Results	36
3.5	Summary	60
CHAPTER 4. CLASSIFICATION ACCURACY ENHANCEMENT THROUGH META-LEARNERS		
4.1	Introduction	61
4.2	Traditional Data Anomaly Based Classification	61
4.3	Graph-Based Medical Data Classification	65
4.4	Rank Based Feature Selection Classification Model With Multivariate Analysis	66
	4.4.1. Data Pre-Processing Algorithm	68
	4.4.2. Fraud Detection Attribute Selection Algorithm	69
4.5	Experimental Results	70
4.6	Summary	76
CHAPTER 5. RE-ENGINEERING OF ENSEMBLES FOR REAL WORLD DATASETS		
5.1	Introduction	77
5.2	Automatic Data Feature Extraction	78
5.3	Traditional Biomedical Data Classification Models	79
5.4	Ensemble Based Partition Classifier Model With Large Number Of Feature Sets	81
	5.4.1 Algorithm	83
5.5	Experimental Results	85
	5.5.1 Anomaly Patterns	92

5.6	Summary	96
CHAPTER 6. CONCLUSIONS AND FUTURESCOPE		
6.1	Conclusion	97
6.2	Future Scope	98
7	REFERENCES	99
8	LIST OF PUBLICATIONS	108