## CONTENTS

<table>
<thead>
<tr>
<th>LIST OF TABLES</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>I</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST OF FIGURES</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>III</td>
</tr>
</tbody>
</table>

<table>
<thead>
<tr>
<th>LIST SYMBOLS AND ABBREVIATIONS</th>
<th>Page No.</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>V</td>
</tr>
</tbody>
</table>

### 1. Introduction

1.1 Motivation

1.2 Data Mining

1.3 Advantages of Data Mining

1.4 Misuses of Data Mining

1.5 What is Information Privacy

1.6 Basic Principles to protect information Privacy

1.7 Privacy preserving data mining

### 2. Privacy Preserving Data Mining- Background Study

2.1 Classification of Datasets

2.1.1 Distributed Database

2.1.1.1 Data partitioning

2.1.2 Centralized Database

2.2 Taxonomy of PPDM

2.3 Randomization

2.4 Group Based Anonymization

2.4.1 The K-anonymity Framework

2.4.2 Personalized Privacy- Preservation

2.4.3 Utility-Based Privacy Preservation

2.4.4 The l-diversity method

2.5 Data Swapping

2.6 Aggregation

2.7 Suppression

2.8 Secure multiparty computation

2.8.1 Cryptography : Oblivious transfer
2.8.2 The Two-party case

2.8.3 The Multi-party case
   2.8.3.1 Trust
   2.8.3.2 Independence
   2.8.3.3 Communication
   2.8.3.4 Privacy
   2.8.3.5 Correctness
   2.8.3.6 Efficiency
   2.8.3.7 Guaranteed Output
   2.8.3.8 Fairness

2.9 Data Transformation

2.10 Information Hiding

2.11 Conclusions

3. Methodology
   3.1 Objectives of the thesis
   3.2 Methodology
   3.3 Steps involved in PPDM using vector quantization
      3.3.1 Proposed Approach 1
      3.3.2 Proposed Approach 2
   3.4 K-mean clustering on transformed data

4. Description of Work
   4.1 Types of Quantization
      4.1.1 Scalar Quantization
      4.1.2 Vector Quantization
   4.2 Design Problem
      4.2.1 Optimality Criteria
      4.2.2 Nearest Neighbor Condition
      4.2.3 Centroid Condition
   4.3 Codebook Generation Algorithms
4.3.1 LBG Algorithm 48
4.3.2 LBG Algorithm Design 49
4.3.3 Codebook generation by LBG 50
4.3.4 Flow chart representation of LBG 51
4.3.5 Codebook by Modified LBG 53
  4.3.5.1 Utility of codeword 53
4.3.6 Flow chart representation of M-lbg 55

5. Experimental Results 57
  5.1 Results 57
  5.2 Performance evaluation 74

6. Conclusions 88

7. References 90

Author Publications 101