# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>CHAPTER</th>
<th>TITLE</th>
<th>PAGE</th>
</tr>
</thead>
<tbody>
<tr>
<td>ACKNOWLEDGMENT</td>
<td></td>
<td>i</td>
</tr>
<tr>
<td>ABSTRACT</td>
<td></td>
<td>iii</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td></td>
<td>vii</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td></td>
<td>ix</td>
</tr>
<tr>
<td>LIST OF ABBREVIATIONS</td>
<td></td>
<td>xi</td>
</tr>
</tbody>
</table>

## 1 INTRODUCTION

1.1 Clustering Terminologies | 2
1.2 Research Problems in Data Clustering | 5
1.3 Scope of the Research | 8
1.4 Motivation of the Research | 9
1.4.1 Google Data Flow | 10
1.4.2 Health Care | 12
1.5 Objectives of the Research | 14
1.6 Research Contributions | 15
1.7 Organization of the Thesis | 16

## 2 REVIEW OF LITERATURE

2.1 Algorithms to Handle the Arrival of Unclustered Data Point
  2.1.1 Adaptive Clustering Algorithms | 22
  2.1.2 Incremental Clustering Algorithms | 24
  2.1.3 Dynamic Clustering Algorithms | 32
2.2 Statistics and Regression Techniques in $k$-means Clustering | 36
3 DISTANCE THRESHOLD ESTIMATION MODEL FOR INTER CLUSTER DATA MIGRATION (ICDM)

3.1 Scenario of ICDM
   3.1.1 Non Occurrence of ICDM
   3.1.2 Occurrence of ICDM

3.2 Influence of Cluster Measures on ICDM
   3.2.1 Cluster Cohesion
   3.2.2 Cluster Separation
   3.2.3 Distance between Centers of Two Closest Clusters

3.3 Significance of Cluster Measures with the Occurrence of ICDM

3.4 Distance Threshold Estimation Model for ICDM

3.5 Experimental Setup and Results
   3.5.1 Synthetic Datasets
   3.5.2 Real Datasets

3.6 Regression Analysis to Find the Occurrence of ICDM
   3.6.1 Multiple Regression Analysis to Estimate Distance Threshold $D$
   3.6.2 Simple Regression Analysis to Estimate the Distance Threshold Interval

3.7 Compactness of the Clustering System
3.8 Distinctiveness of the Clustering System

3.9 Summary