Contents

1. Introduction 1

1.1 Privacy Preserving Data Mining (PPDM) 1

1.2 Classification of PPDM 4

1.3 Privacy Preserving Technique 5

1.4 Privacy Preserving Clustering 9

1.5 Data mining and Privacy 9

1.6 Applications of PPDM 13

1.7 Aim of the Work 13

1.8 Report Organization 13

2. Microaggregation 14

2.1 Statistical Databases 14

2.2 Statistical Disclosure Control 14

2.3 Micraggregation Techniques (MATs) 15

2.3.1 Uni-variate MATs 17

2.3.2 Data-oriented Microaggregation 18

2.4 Multi-variate MATs 18

2.5 Optimal Microaggregation 19

2.6 Issues Related to MATs 19

2.7 Clustering versus MATs 22

3. Related Works 23

4. Background of the Work 29

4.1 Clustering 29

4.2 Challenges in Clustering 29

4.3 Categorization of Clustering 30
4.4 A Density-Based Notion of Clusters 32
4.5 Cluster with respect to Eps and MinPts 35
4.6 Noise 35
4.7 DBCSAN 35
4.8 MDAV 37

5. Proposed Method 38

6. Experimental Results 41
6.1 Overview of Datasets 41
6.2 Quantification of Privacy 42
6.3 Performance Analysis 44

7. Conclusions and Future Work 48

8. References 49