## Contents

1 Introduction 1

1.1 Background 1

1.2 Motivation 2

1.3 Contributions 4

1.4 Organization of the Thesis 6

2 Problem Definition and Related Work 8

2.1 Subspace Clustering Problem 8

2.2 CLIQUE algorithm 12

2.3 Improvements over CLIQUE 17

3 AOSS: An Attribute Oriented Storage Structure 21

3.1 Limitations of the existing storage techniques. 22

3.2 Design of Attribute Oriented Storage Structure 24

3.3 Database operations Using AOSS 28

3.4 Efficiency obtained using AOSS 34

3.5 Experimental Results 40

3.5.1 Synthetic data generation 41

3.5.2 Synthetic data results 42

3.6 Summary 45

4 SAMCLIQ: A SAMpling based CLIQue algorithm 47

4.1 Use of Sampling in Data Mining 48
4.1.1 Role played by sampling in data mining
4.1.2 Limitations of sampling

4.2 Proposed Sampling technique
4.2.1 Criteria for a good sample
4.2.2 Sampling for finding frequent sets
4.2.3 AOSS based sampling technique

4.3 Subspace Clustering Using Sampling
4.3.1 Algorithm for identification of dense units

4.4 Experimental Results
4.4.1 Synthetic data generation
4.4.2 Synthetic data results

4.5 Summary

5 MLSCLUS: A Multi Level Subspace CLUSTERing Algorithm
5.1 Use of Maximal Frequent Itemsets in Subspace Clustering
5.2 MADUGEN: A Maximal Dense Unit Generation Algorithm Using Multiple Threshold Values
5.2.1 Experimental results
5.2.2 MADUGENMT: MADUGEN algorithm with multiple threshold values

5.3 AOMLSCLUS: An Attribute Oriented Multi Level Subspace CLUSTERing Algorithm
5.3.1 Experimental results
5.4 Summary

6 Discussion

6.1 Characteristics of the AOSS method

6.2 Extensions and Applications of Subspace Clustering methods

7 Conclusions

7.1 Summary of The Thesis

7.2 Future Research Directions

References