# CONTENTS

1. Introduction

1.1 Data Mining \hspace{1cm} 1
1.2 Evolution of Database System Technology \hspace{1cm} 1
1.3 Knowledge Discovery in Databases (KDD) \hspace{1cm} 5
1.4 Architecture of Data Mining System \hspace{1cm} 7
1.5 Data Mining on different kinds of Data \hspace{1cm} 11
1.6 Data Mining Process \hspace{1cm} 12
1.7 Application Areas of Data Mining \hspace{1cm} 15
1.8 Classification of Data Mining System \hspace{1cm} 17
1.9 Data Mining Functionalities \hspace{1cm} 20

1.9.1 Concept/Class Description: characterization and discrimination \hspace{1cm} 21
1.9.2 Mining Frequent Patterns, Associations and Correlations \hspace{1cm} 22
1.9.3 Classification and Prediction \hspace{1cm} 24
1.9.4 Cluster Analysis \hspace{1cm} 25
1.9.5 Outlier Analysis \hspace{1cm} 26
1.9.6 Evolution Analysis \hspace{1cm} 26

1.10 Frequent Itemsets and Association Rules \hspace{1cm} 27
1.11 Organization of the thesis \hspace{1cm} 36

2. Literature survey

2.1 Introduction \hspace{1cm} 37
2.2 Survey on Frequent itemset Mining Algorithms \hspace{1cm} 38
2.3 Survey on Association Rule Measures \hspace{1cm} 43
2.4 Problem description and objectives \hspace{1cm} 46

3. SMine Frequent Itemset Mining Algorithm

3.1 Introduction \hspace{1cm} 50

3.1.1 Nominal Attributes \hspace{1cm} 50
3.1.2 Data structures \hspace{1cm} 51
3.2 Steps in SMine Algorithm \hspace{1cm} 52
3.3 SMine Algorithm \hspace{1cm} 55

3.3.1 Algorithm \hspace{1cm} 55
3.3.2 Explanation \hspace{1cm} 58
3.4 Example \hspace{1cm} 60
3.5 Results and Discussions
   3.5.1 Weather Nominal Dataset [W2] 64
   3.5.2 Hepatitis Dataset [W3] 70
   3.5.3 Mushroom Dataset [W3] 77

3.6 Summary 83

4. Modified SMine Frequent Itemset Mining Algorithm

4.1 Introduction 85
   4.1.1 Nominal Attributes 85
   4.1.2 Data structures 85

4.2 Steps in Modified SMine Algorithm 87
4.3 Modified SMine Algorithm 89
   4.3.1 Algorithm 89
   4.3.2 Explanation 92

4.4 Example 94

4.5 Results and Discussions 98
   4.5.1 Weather Nominal Dataset [W2] 99
   4.5.2 Hepatitis Dataset [W3] 102
   4.5.3 Mushroom Dataset [W3] 107

4.6 Handling Large Databases 110
   4.6.1 Partition Algorithm 114
   4.6.2 Results and Discussions 116

4.7 Summary 120

5. Novel Measures in the Generation of Association Rules

5.1 Introduction 123
5.2 Proposed Measures 124
   5.2.1 Modified-Support 124
   5.2.2 Modified-Confidence 129

5.3 Modified-Support and Modified-Confidence 135
   5.3.1 Algorithm 135
   5.3.2 Explanation 136

5.4 Example 136
5.5 Results and Discussions 143
   5.5.1 Weather Nominal Dataset [W2] 143
      5.5.1.1 Rules Generated for Case 1 144
      5.5.1.2 Rules Generated for Case 2 148
      5.5.1.3 Rules Generated for Case 3 149
      5.5.1.4 Comparison of Rules Generation 150
5.5.2 Hepatitis Dataset [W3] 151
  5.5.2.1 Rules Generated for Case 1 152
  5.5.2.2 Rules Generated for Case 2 159
  5.5.2.3 Rules Generated for Case 3 163
  5.5.2.4 Rules Generated for Case 4 164
  5.5.2.5 Comparison of Rules Generation 166

5.5.3 Mushroom Dataset [W3] 167
  5.5.3.1 Rules Generated for Case 1 168
  5.5.3.2 Rules Generated for Case 2 172
  5.5.3.3 Rules Generated for Case 3 176
  5.5.3.4 Comparison of Rules Generation 177

5.6 Time Complexity 179
5.7 Summary 181

6. Conclusion and Future work
  6.1 Conclusion 182
  6.2 Future Work 183
    Publications 184
    References 186

Appendix