

# Index

---

<b>Chapter 1 : Basic Concepts</b>	<b>1-1 to 1-18</b>
1.1 Big Data.....	1-2
1.1.1 What is Big Data ?.....	1-2
1.1.2 Need of Big Data.....	1-3
1.2 Data mining and Knowledge Discovery.....	1-5
1.3 Data mining Architecture.....	1-5
1.3.1 The Origins of Data Mining.....	1-7
1.4 Data Mining In Business Analytics.....	1-9
1.4.1 How does Data Mining Inform “Business Analytics”?.....	1-9
1.4.2 Data Mining Techniques in “Business Analytics”.....	1-9
1.5 Data mining Tasks.....	1-10
1.6 Data mining Process.....	1-13
1.6.1 Types of Data mining Processes.....	1-13
1.7 Real World Data Mining Application.....	1-15
<b>Chapter 2 : Data and Pre-Processing</b>	<b>2-1 to 2-26</b>
2.1 Understanding of Data, what is Data?.....	2-2
2.1.1 The Type of Data.....	2-2
2.2 The Type of an Attribute.....	2-3
2.2.1 Describing Attributes by the Number of Values.....	2-5
2.3 Types of Data Sets.....	2-6
2.3.1 Graph-based data.....	2-8
2.3.2 Ordered data.....	2-9
2.4 Data Quality.....	2-12
2.4.1 Measurement and Data Collection Issues.....	2-12
2.4.2 Measurement and Data Collection Errors.....	2-12
2.4.3 Noise and Artifacts.....	2-13
2.4.4 Precision, Bias, and Accuracy.....	2-13
2.4.5 Inconsistent Values.....	2-15
2.4.6 Duplicate Data.....	2-16
2.5 Sampling.....	2-18
2.5.1 Types of Sampling.....	2-18
2.5.2 Sampling Approaches.....	2-19

---



2.5.3	Progressive Sampling.....	2-20
2.6	Data Normalization.....	2-20
2.7	Data Cleaning.....	2-21
2.8	Similarity Measures.....	2-24
2.9	Feature Subset Selection.....	2-25

---

**Chapter 3 : Classification****3-1 to 3-8**

3.1	Decision-Tree Based Approach.....	3-2
3.2	Rule-based Approach.....	3-3
3.3	Instance-based Classifiers.....	3-3
3.4	Support Vectors Machine.....	3-4
3.5	Ensemble Learning.....	3-5
3.6	Applications : B2B Customer Buying Stage Prediction, Recommender Systems.....	3-8

---

**Chapter 4 : Clustering****4-1 to 4-11**

4.1	Introduction.....	4-2
4.1.1	Clustering Methods.....	4-2
4.2	Hierarchical clustering Methods.....	4-3
4.3	Partitional Vs Hierarchical clustering.....	4-7
4.4	Graph based methods.....	4-7
4.5	Density Based Methods.....	4-8
4.6	Cluster validation.....	4-10
4.7	Applications : Customer Profiling, Market segmentation.....	4-10

---

**Chapter 5 : Association Analysis****5-1 to 5-16**

5.1	Apriori Algorithm and its Extensions.....	5-2
5.2	Association Pattern Evaluation.....	5-3
5.3	Sequential Patterns Mining.....	5-4
5.4	Frequent Sub graph Mining.....	5-6
5.5	Applications.....	5-10
5.6	Anomaly Detection.....	5-14
5.7	Ethics of Data Mining, Privacy, what can/do Firms know?.....	5-16