Sarvajanik College of Engineering & Technology
Information Technology Department
Semester: 7
Data Mining and Business Intelligence
Subcode: 2170715

Tutorial 1

No.	Question
1.	Define the term Data Mining. Explain why Data Mining is crucial for business intelligence.
2.	What is KDD? Explain the steps in KDD with the help of a neat diagram.
3.	Explain the reasons for noise in data. What are the ways to deal with noisy data.
4.	Explain the reasons for missing values in data. What are the ways to deal with missing data.
5.	What is data normalization? Explain the following methods with suitable example: (i) min-max normalization, (ii) z-score normalization, (iii) decimal scaling
6.	Briefly explain issues in data mining.
7.	What is an outlier? How can outliers be detected?
8.	What is a data warehouse? With the help of a neat diagram explain its 3-tier architecture.
9.	Explain metadata in a data warehouse.
10.	What is meant by data mart?
11.	Give the differences between OLAP and OLTP.
12.	Explain the following OLAP operations with suitable example: Roll up, Drill down, slice, dice and pivot.
13.	Consider the PlayTennis dataset given below:

Outlook	Temperature	Humidity	Wind	PlayTennis
Sunny	Hot	High	Weak	No
Sunny	Hot	High	Strong	No
Overcast	Hot	High	Weak	Yes
Rain	Mild	High	Weak	Yes
Rain	Cool	Normal	Weak	Yes
Rain	Cool	Normal	Strong	No
Overcast	Cool	Normal	Strong	Yes
Sunny	Mild	High	Weak	No
Sunny	Cool	Normal	Weak	Yes
Rain	Mild	Normal	Weak	Yes
Sunny	Mild	Normal	Strong	Yes
Overcast	Mild	High	Strong	Yes
Overcast	Hot	Normal	Weak	Yes
Rain	Mild	High	Strong	No

Now classify the following data sample using naïve Bayesian classifier:

X= (Outlook="Sunny", Temperature="Cool", Humidity="High", Wind="Strong")

14.	Explain the steps of the ID3 algorithm for Decision Tree Generation.
15.	Explain the statistical terms: (i) Entropy and (ii) Information Gain.
	Generate a decision tree from the PlayTennis dataset using algorithm ID3.
16.	Write a short note on "Web Mining".
17.	Briefly explain the following data warehouse schemas with suitable illustration: (i) Star Schema, (ii) Snowflake Schema, (iii) Fact Constellation Schema.