JSPM's
Bhivarabai Sawant Institute of Technology & Research
DEPARTMENT OF COMPUTER ENGINEEREING
UNIT TEST 1
Subject: Elective I(Data Mining and Warehousing)

Time:1 Hour **Max Marks:30**

Instruction:

- **Solve 1or 2, 3 or 4, 5 or 6.**
- **Draw neat diagram wherever necessary.**
- **Figures to right indicate full marks.**

Q1.a) What is data pre-processing? Explain the different steps in data Preprocessing **6 Marks**

b) Compare OLTP and OLAP **4Marks**

Or

Q2. a) Consider the following group of data **6 Marks**

200, 300, 400, 600,1000

i) Use the min-max normalization to transform value 600 onto the range [0.0,1.0].

ii) Use the decimal scaling to transform value 600.

b) In real-world data, tuples with missing values for some attributes are a common occurrence. Describe various methods for handling this problem **4Marks**

Q3. a) Describe data warehouse architecture with diagram. **6 Marks**

b) Explain Data cube multidimensional data model. **4Marks**

or

Q4.a) Explain Following with the help of example and draw the diagram: **6 Marks**

i)Stars Schema

ii) Snowflakes Schemas

iii) Fact Constellations

b) What are the difference between the three main type of Data warehouse usage: Information processing, Analytical processing, Data mining **4Marks**

Q5a).Two objects represented by tuples (22,1,42,10) and (20,0,36,8) **6 Marks**

a) Compute the Euclidean distance between the two objects

b) Compute the Manhattan distance

c) Compute Minskowski Distance

b) Explain Data matrix versus Dissimilarity matrix **4Marks**

Or

Q6a) .Explain the following terms:- **6 Marks**

a) Minskowski Distance

b) Euclidean Distance

C) Manhattan Distance

b) Explain Cosine similarity in brief **4Marks**