Dr. Jayashree

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

T-2 EXAMINATION, April 2018

B.Tech (BI) VI Semester

COURSE CODE: 16B11BI112 **MAX. MARKS: 25**

COURSE NAME: Datawarehousing and Mining for Bioinformatics

COURSE CREDITS: 04 **MAX. TIME: 1.5HR**

Note: All questions are compulsory. Carrying of mobile phone during examinations will be treated as case of unfair means. Marks are indicated against each question.

---

1. Describe the application of identification trees for the prediction of (5)
   - (i) Protein secondary structure
   - (ii) HIV protease cleavage site

2. Which criteria are used for constructing identification trees? Compare these criteria in terms of their relative significance. (2)

3. Apply the average linkage and complete linkage (furthest neighbor) methods to the following dataset. Compare the results. (8)

Person	X_1	X_2
a	2	4
b	8	2
c	9	3
d	1	5
e	8.5	1

4. Explain the limitations of using a single identification tree. How can these be circumvented by combining multiple identification trees? (5)

5. Why is the OPTICS method of clustering better than the DBSCAN method? Elaborate. (2)

6. Differentiate between k-medoids, CLARA and CLARANS methods. (3)