JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST-2 EXAMINATION- April, 2023

B.Tech. (BI) VIth Semester

COURSE CODE: 18B1WBI632 (3) MAX. MARKS: 25

COURSE NAME: Dataware Housing and Mining for Bioinformatics

COURSE INSTRUCTORS: Dr. Ekta Gandotra MAX. TIME: 1.5 Hrs.

Note: *All questions are compulsory. Marks are indicated against each question in square brackets.*

Q1. a. What does a subject-oriented data warehouse signify? Give an example. [2] CO1

b. How do you know if your model is overfitting or underfitting? Explain with the help of a suitable example. [3] CO3

Q2. Find the class of the test sample using k-NN algorithm for the following dataset. Take k=3 and k=5. Use L2 Norm for distance computations. [4] CO4

Brightness	Saturation	Class
40	20	Red
50	50	Blue
60	90	Blue
10	25	Red
70	70	Blue
60	10	Red
25	80	Blue

Test Sample.

Brightness	Saturation	Class
20	35	?

Q3. How do you determine the value of k in the k-NN algorithm? What are the drawbacks of choosing a value of k that is either too small or too large? [3] CO4

Q4. Give any three drawbacks of ID3 Decision Tree algorithm. Suggest methods to overcome these drawbacks. [3] CO4

Q5. Consider the following training data which predicts if the students get **passed** in a course (Yes, No), based on their previous **CGPA** (High, Medium, Low) and whether they **studied** or not. CO5

i. Compute the Gini index for the overall collection of training examples. [2]

ii. Compute the Gini index for the CGPA attribute. [2]

iii. Compute the overall entropy of the training examples. [2]

iv. Find the root node of the decision tree using ID3 algorithm. [4]

Example	1	2	3	4	5	6
CGPA	L	L	M	M	H	H
Studied	F	T	F	T	F	T
Passed	No	Yes	No	Yes	Yes	Yes

Note: Show all steps of calculations.