

JAYPEE UNIVERSITY OF INFORMATION TECHNOLOGY, WAKNAGHAT

TEST -1 EXAMINATION- February, 2020

B.Tech. (CSE, IT) VI Semester

COURSE CODE: 18B1WCI634

MAX. MARKS: 15

COURSE NAME: Machine Learning

COURSE CREDITS: 2

MAX. TIME: 1 Hr.

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

- Q1. CO-3 a. What is ROC curve? Give its significance. [2]
b. For a multiclass problem, the number of instances of three classes A, B and C are 100, 150, and 50 respectively. Calculate the value of accuracy, precision and recall from the following confusion matrix. [3]

		Predicted→		
		A	B	C
Actual	A	80	10	10
	B	10	120	20
	C	0	5	45

- Q2. CO-4 Consider the following dataset which predicts if the students **pass** machine learning course (Yes, No), based on their previous **GPA** (High, Medium, Low) and whether or not they **Studied**. Use $\log_2 3 = 1.6$. [1]
a. Find the entropy $H(\text{Passed})$? [1]
b. Find the entropy of attributes i.e. $H(\text{Passed}|\text{GPA})$ and $H(\text{Passed}|\text{Studied})$ [2]
c. Draw a full decision tree that would be learned for this dataset. [2]

Instance	1	2	3	4	5	6
GPA	L	L	M	M	H	H
Studied	F	T	F	T	F	T
Passed	No	Yes	No	Yes	Yes	Yes

- Q3. CO-2 a. What is overfitting and underfitting in machine learning. Express in terms of bias and variance. [2]
b. Using the following dataset, predict the class for the record (Color=Red, Type=SUV, Origin=Domestic) using Naïve Bayes algorithm. [3]

Instance	Color	Type	Origin	Stolen?
1	Red	Sports	Domestic	Yes
2	Red	Sports	Domestic	No
3	Red	Sports	Domestic	Yes
4	Yellow	Sports	Domestic	No
5	Yellow	Sports	Imported	Yes
6	Yellow	SUV	Imported	No
7	Yellow	SUV	Imported	Yes
8	Yellow	SUV	Domestic	No
9	Red	SUV	Imported	No
10	Red	Sports	Imported	Yes