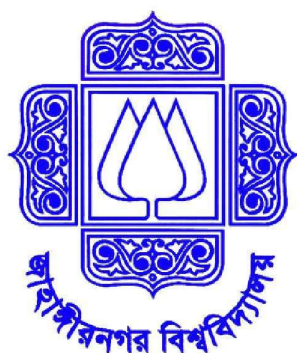


# JAHANGIRNAGAR UNIVERSITY

Institute of Information Technology (IIT)



## CLASS TEST 1

Summer Semester 2023 Intake

Date of Examination: JULY 14, 2023

---

**HASAN TAHSIN RAFSAN**

Professional Masters in Information Technology (PMIT) Program

1<sup>ST</sup> Trimester Regular Batch

Section A

ID No: 232137

232437

P-1

Ans: 1

So, in the question there are 4 activities.

Prob (a) is a data mining task. So answer is "yes"

It belongs to "Classification" because here the analyst does not have any ~~pre~~ prior information about ~~these~~ objects. So, he need to predict these ~~prob~~ based on given models.

prob (b) is a data mining tasks also. ~~So~~ So answer is "yes". This belongs to "Clustering". Here, the doctor has to ~~lab~~ labelled the CT scan reports to make result.

prob (c) is also a data mining task. So answer is "yes". This ~~lab~~ belongs to "Regression".

Because, here they have to predict the value based on given present variable based values & models.

p-2

232438

Prob (d) is also a data mining task, so the answer is "Yes". It belongs to "Anomaly Detection". Here, monitoring heart rate of patient abnormalities & the terms of unusual patterns of diseases are quite same. The characteristics are different from observed data.

Ans: B

3 Data quality problems are (i) noise (ii) Missing values. (iii) Duplicate data.

For noise it is an extraneous object where some attributes at- tribute to the modification of original values. For example: for voice, there might be magnitude & shape of original signal is destroyed based on the waves combined & temperature.

232437

1-3

Missing values. generally created for information not collected. or attribute may not be applicable for all cases.

{ i.e. - people decline to tell their age, weights or children doesn't applicable for income or handling estimating values.

Another one is duplicate data where many people ~~to~~ ~~change~~ sometimes ~~re~~ receive duplicate mailings because they appear in a database ~~to~~ multiple times under different names. ~~one to other changed~~

This ~~data~~ duplicate data or multiple sets of data causes the issues.

232437

P-4

Ans. 5

Temperature can either be an interval or ratio measurement statistically. If one measure temperature in degrees ~~on~~ Fahrenheit or Celsius, is considered to interval. ~~because two points~~ zero points are arbitrary. This can be below zero degrees. Fahrenheit or Celsius.

On the other hand, if one measure temperature in Kelvin, then it will be ratio as the zero point is absolute. There cannot be temperature below zero Kelvin.

232137

Ans: 2

Credit card fraud detection :- A credit card company records the transactions made by every credit card holder, along with personal info such as limit, age, income & address. When a new transaction arrives, it's compared against the profile of the user. If the characteristics of the transaction are very different from previously created profile, then the transaction is flagged as potentially fraudulent.