

Chapter 1

1. Discuss whether or not each of the following activities is a data mining task.

Answer : b e f g

2. For each of the following data sets, explain whether or not data privacy is an important issue.

Answer :

- (a) Not important, because this is public data
- (b) Not important, because this is public data
- (c) Not important, because this is public data
- (d) Not important, because this is public data
- (e) Not important, because this is public data

Chapter 2

2. Classify the following attributes as binary, discrete, or continuous. Also classify them as qualitative (nominal or ordinal) or quantitative (interval or ratio). Some cases may have more than one interpretation, so briefly indicate your reasoning if you think there may be some ambiguity.

Answer :

- (a) Time in terms of AM or PM.

Binary, qualitative, nominal

- (b) Brightness as measured by a light meter.

Continuous, quantitative, ratio

- (c) Brightness as measured by people's judgments.

Discrete, qualitative, ordinal

- (d) Angles as measured in degrees between 0° and 360°.

Continuous, quantitative, ratio

- (e) Bronze, Silver, and Gold medals as awarded at the Olympics.

Discrete, qualitative, ordinal

- (f) Height above sea level.

Continuous, quantitative, interval/ratio

(g) Number of patients in a hospital.

Discrete, quantitative, ratio

(h) ISBN numbers for books. (Look up the format on the Web.)

Discrete, qualitative, nominal

(i) Ability to pass light in terms of the following values: opaque, translucent, transparent.

Discrete, qualitative, ordinal

(j) Military rank.

Discrete, qualitative, ordinal

(k) Distance from the center of campus.

Continuous, quantitative, interval/ratio (depends)

(l) Density of a substance in grams per cubic centimeter.

Continuous, quantitative, ratio

(m) Coat check number. (When you attend an event, you can often give your coat to someone who, in turn, gives you a number that you can use to claim your coat when you leave.)

Discrete, qualitative, nominal

3.a Who is right, the marketing director or his boss? If you answered, his boss, what would you do to fix the measure of satisfaction?

Answer :

Clearly boss is right here. And apparently the marketing director missed a very obvious thing i.e. The total number of people using the product. Since, the product which is being used by many people will have a lot of complaints as well. So a better measure of satisfaction would be:

Number of complaints for the product / Total number of people using that product X

3.b What can you say about the attribute type of the original product satisfaction attribute

Answer :

type of the original product satisfaction attribute are not ratio attributes, are Ordinal attributes

7. Which of the following quantities is likely to show more temporal autocorrelation: daily rainfall or daily temperature? Why?

Answer :

daily temperature, Because the temperature of the day changes with time

12. Distinguish between noise and outliers. Be sure to consider the following questions.

Answer :

a. Is noise ever interesting or desirable? Outliers?

no, outliers is

b. Can noise objects be outliers?

yes

c. Are noise objects always outliers?

no

d. Are outliers always noise objects?

no

e. Can noise make a typical value into an unusual one, or vice versa?

yes, no