Solution Student Handbook – Chapter 2

Objective Type Questions

Please choose the correct option in the questions below:

1.A school named ABC has recorded the total marks of every student in the class. This is an example of :

   a.   Qualitative data
   b.   Quantitative data
   c.   Both Qualitative and Quantitative data
   d.   None of the above.

2.A food delivery app has asked for your feedback on the quality of the food. You have written two paragraphs to describe the food. This is an example of:

   a.   Qualitative data
   b.   Quantitative data
   c.   Both Qualitative and Quantitative data
   d.   None of the above

3.It would help if you predicted what the temperature would be for next Friday.

   a.   Clustering
   b.   Regression
   c.   Anomaly detection
   d.   Binary classification

4.You need to predict if your car tire will last for the next 1000 km. Which algorithm will you use?

   a.   Clustering
   b.   Regression
   c.   Anomaly detection
   d.   Binary classification

5.You want to build a way to segregate spam emails from good emails. Which algorithm will you use?

   a.   Clustering
   b.   Regression
   c.   Anomaly detection
   d.   Binary classification

Standard Questions

Please answer the questions below in no less than 100 words:

(1) What are the common career paths for data science?

Data Scientist, Business Intelligence Analyst, Data Mining Engineer, Data Architect, Senior Data Scientist etc are some of the common career paths for data science.

(2) What does Data Architect do?

Data Architects are the person who works together with the users or with the designers or developers of the system in order to make a blue print which is used by the data management systems for centralizing, integrating and maintaining of the data sources.

(3) What are the difference between classification and regression?

The difference between classification and regression are as followed:

a. **Classification**: Classification helps us in order to predict that any new item which falls or belongs to any of the category/class A or B. If the person tries to make the prediction between more than two options, then it is termed as multiclass algorithms.
b. **Regression**: The method of predicting the value of a continuous variable is to be termed as the Regression.

High Order Thinking Skills (HOTS)

Please answer the questions below in no less than 200 words:

(1) Discuss a recent innovation that makes use of reinforcement learning.

Sometimes the problems related to the models of automatic robots or self driving cars occurs. Such problems needs to be solved by making decisions which should be tally based on the external factors. Hence reinforcement learning helps in solving such type of occurred problems of the models by taking decision on the external factors.

(2) Write a short note on how data science is helping sports teams.

Data science helps the sports teams by making the future predictive analysis which will help the teams in overcoming/fulfilling all the necessary things which is to be required to be fulfilled during the game day. Nowadays some sports channels makes use of data science in order to predict the overall performance of the team or players in the game or tournament. It also helps in making accurate decisions.

Applied Projects

Emails are a part of daily communication. Sometimes we receive unwanted emails called spam. There are few techniques that email providers use to identify spam emails.

Content base filtering (Analysing the words, occurrence, distribution of words to identify spam mail)

Header filters (Reviewing the email header) Example: Promo, Offer!

Provide 2 examples each of words /phrases in email content & header which marks an email as spam. Explain in detail, how email providers make use of clustering to mark an email as spam. Also elaborate how email providers create and update the words/phrases to mark an email as spam.