ANSWERS

ANSWER 3

"A chemist wants to find some interesting patterns in which patients are behaving upon administering the drug"

As data scientists, we have to classify our problem statements into four problem categories, which are:

1. Prediction/Forecasting
2. Classification
3. Optimization
4. Unsupervised Learning

Now, we aren't predicting how the patients will behave when they are given a specific drug, since we don't have any historical data on how the patients have reacted to that same drug previously.

We don't have specific categories in which we'll be putting the patients according to the effects observed when they take in the drug hence, we can't use classification technique either.

Our problem statement doesn't demand from us to improve/ work on the side effects or the effects of the drug administered by the patients, although it might be the consequence of resulting patterns of the problem statement. But since we aren't there yet, it doesn't come under the ambit of an optimization problem.

The problem statement wants us to find the patterns of behavior after the patients administer the drug and then classify them in sets of those patterns. We don't have any historical data nor any classification categories, or any optimization problem. Here, we will be learning as we conduct our experiment/research hence, the problem category for the given problem statement would be unsupervised learning.

ANSWER 4

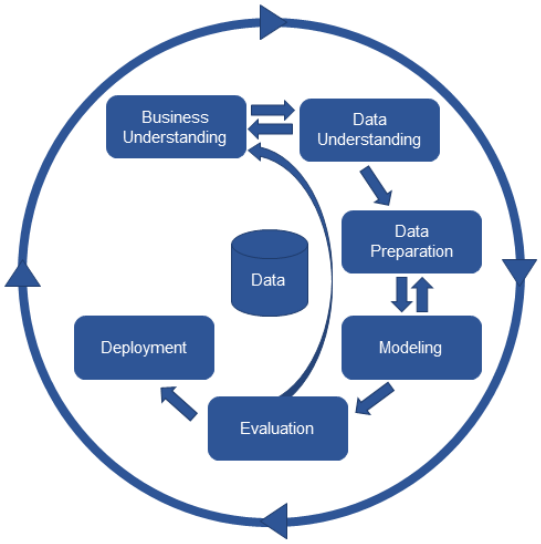
In order to select a suitable machine learning algorithm for a given problem statement, as a data scientist we have to carry out an initial analysis and then go with the most preferred data mining process and choosing the most suitable machine learning model.

CRISP-DM(CRoss Industry Standard Process for Data Mining) is a structure we use to perform the initial step of analysis and decide which model will be the most convenient to solve a business problem. It describes the commonly used approach to solve a business problem.

With CRISP-DM's architecture, the procedure to select a suitable machine learning algorithm for a problem statement will be carried out in following steps:

1. Understand client business and define the business objective out of the problem statement.
2. Understand the data which includes asking the client for the data and exploring the raw data which can either be primary or secondary.
3. After obtaining the raw data, we will organize it properly and prepare it in a manner in which it can be studied by us to analyze it properly. These steps are the most important ones and will require 80% of our efforts.
4. The previous step will be followed by modelling in which we choose the machine learning model after successfully analyzing the data present to us.
5. It will lead to the evaluation of the model being used by us, and will be checked if it satisfies the objective it set out to accomplish.
6. Finally, if the evaluation is a success, then we will deploy the model. If not, then we will go back to the previous steps to optimize our model.

This will be better explained with the help of a flowchart:



ANSWER 5

Education industry is a very adaptive and at the same time very evolving field, given that we encounter disrupting technologies, innovations and discoveries every year, and to develop a skilled youth, we need to keep changing our education system to cater the needs of the future. Be it automation, blockchain, sustainability, in order to ensure the growth of our society we need to make sure we are learning new things every day.

Hence, we need to find what curriculum will be most suitable for students that help them develop into skilled human resource. And living in such an evolving atmosphere, we will have to keep changing the curriculum in every 5 years. Hence, we should be able to decide on the most suitable curriculum for them for the next 5 years, so that the knowledge gained doesn't turn out to be useless.