

## : Problem - to - Approach :

### 1. Select the correct statement.

Ans: A methodology is a system of methods used in a particular area of study or activity.

### 2. The first stage of the data science methodology is Data Understanding.

Ans: False.

### 3. Business Understanding is an important stage in the data science methodology. Why?

Ans:

Because it involves domain expertise.

Because it clearly defines the problem and the needs from a business perspective.

Because it ensures that the work generates the intended solution.

Because it shapes the rest of the methodological steps.

### 4. Which of the following statements about the analytic approach are correct?

Ans:

If the question defined in the business understanding deals with exploring relationships between different factors, then a descriptive approach, where clusters of similar activities based on events and preferences are examined, would be the right analytic method.

If the question defined in the business understanding stage can be answered by determining probabilities of an action, then a predictive model would be the right analytic approach.

### 5. For the case study, a decision tree classification model was used to identify the combination of conditions leading to each patient's outcome.

Ans: True.

## : Requirements - to - Collection :

### 1. Which of the following analogies is used in the videos to explain the Data Requirements and Data Collection stages of the data science methodology?

Ans: You can think of the Data Requirements and Data Collection stages as a cooking task, where the problem at hand is a recipe, and the data to answer the question is the ingredients.

### 2. The Data Requirements stage of the data science methodology involves identifying the necessary data content, formats and sources for initial data collection.

Ans: True.

### 3. Which of the following statements are correct?

Ans:

Data scientists determine how to collect the data.

Data scientists identify the data that is required for data modeling.

Data scientists determine how to prepare the data.

### 4. In the Data Collection stage, the business understanding of the problem is revised and decisions are made as to whether or not more data is needed.

Ans: False.

### 5. In the Data Collection stage, techniques such as descriptive statistics and visualization can be applied to the data set, to assess the content, quality, and initial insights about the data

Ans: True.

## : Modeling - to - Evaluation :

### 1. Select the correct statement.

Ans: A training set is used for predictive modeling.

### 2. A statistician calls a false-negative, a type I error, and a false-positive, a type II error.

Ans: False

### 3. The Modeling stage is followed by the Analytic Approach stage.

Ans: False.

### 4. Model Evaluation includes ensuring that the data are properly handled and interpreted.

Ans: True.

### 5. Select the correct statements about the ROC curve?

Ans:

The ROC curve is a useful diagnostic tool for determining the optimal classification model.

By plotting the true-positive rate against the false-positive rate for different values of the relative misclassification cost, the ROC curve can be used to select the optimal model.

ROC stands for Receiver Operating Characteristic curve, which was originally developed to detect enemy air-crafts on radar.

## : Understanding - to - Preparation :

### 1. The Data Understanding stage refers to the stage of removing redundant data.

Ans: True.

### 2. In the case study, working through the Data Understanding stage, it was revealed that the initial definition was not capturing all of the congestive heart failure admissions that were expected, based on clinical experience.

Ans: True.

### 3. The Data Preparation stage encompasses all activities related to constructing the data set.

Ans: False.

### 4. Select the correct statement about what data scientists and database administrators (DBAs) do during the Data Preparation stage.

Ans:

During the Data Preparation stage, data scientists and DBAs define the variables to be used in the model.

During the Data Preparation stage, data scientists and DBAs determine the timing of events.

During the Data Preparation stage, data scientists and DBAs aggregate the data and merge them from different sources.

During the Data Preparation stage, data scientists and DBAs identify missing data.

All of the above statements are correct.

**5. The Data Preparation stage is a very iterative and complicated stage that cannot be accelerated through automation.**

Ans: False.

**: Deployment - to - Feedback :**

**1. The final stages of the data science methodology are an iterative cycle between Modelling, Evaluation, Deployment, and Feedback.**

Ans: True

**2. Feedback is not required once the model is deployed because the Model Evaluation stage would have assessed the model and made sure that it performed well.**

Ans: False.

**3. Deploying a model into production represents the end of the iterative process that includes Feedback, Model Refinement, and Redeployment.**

Ans: True (wrong)

**4. The data science methodology is a specific strategy that guides processes and activities relating to data science only for text analytics.**

Ans: False.

**5. A data scientist determines that building a recommender system is the solution for a particular business problem at hand. What stage of the data science methodology does this represent?**

Ans: Analytic Approach.

**6. A car company asked a data scientist to determine what type of customers are more likely to purchase their vehicles. However, the data comes from several sources and is in a relatively “raw format”. What kind of processing can the data scientist perform on the data to prepare it for the Modeling stage?**

Ans:

Addressing missing/invalid values.

Feature engineering.

Combining the data from the various sources.

Transforming the data into more useful variables.

**7. What do data scientists typically use for exploratory analysis of data and to get acquainted with it?**

Ans: They use descriptive statistics and data visualization techniques.

**8. Data scientists may frequently return to a previous stage to make adjustments, as they learn more about the data and the modeling.**

Ans: True.

**9. For predictive models, a test set, which is similar to – but independent of – the training set, is used to determine how well the model predicts outcomes. This is an example of what step in the methodology?**

Ans: Model Evaluation.

**10. Why should data scientists maintain continuous communication with business sponsors throughout a project?**

Ans:

So that business sponsors can review intermediate findings.

So that business sponsors can ensure the work remains on track to generate the intended solution.

So that business sponsors can provide domain expertise.