Course 3- Data Science Methodology

Week 2

Quiz 1: From Understanding to Preparation

1.

Question 1

In the case study, working through the **Data Preparation** 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.

1 / 1 point

○

True

◉

False

Correct

Correct. It was working through the **Data Understanding** staging that the initial definition was found to be incomplete.

2.

Question 2

Select the correct statement about what data scientists do during the Data Preparation stage.

1 / 1 point

○

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

○

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

○

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

○

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

◉

All of the above statements are correct.

Correct

Correct.

3.

Question 3

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

1 / 1 point

○

True

◉

False

Correct

Correct.

Quiz 2: From Modeling to Evaluation

1.

Question 1

The Modeling stage is followed by the Analytic Approach stage.

1 / 1 point

○

True

◉

False

Correct

Correct.

2.

Question 2

Select the correct statement(s) about the Model Evaluation stage of the data science methodology.

1 / 1 point

☐

Model Evaluation cannot include statistical significance testing.

☑

Model Evaluation includes ensuring the model is designed as intended.

Correct

Correct.

☑

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

Correct

Correct.

☑

Model Evaluation includes ensuring that the model is working as intended.

Correct

Correct.

3.

Question 3

Select the correct statements about the ROC curve.

1 / 1 point

☑

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.

Correct

Correct.

☐

The ROC curve was originally developed to optimize healthcare and detect congestive heart failure readmission rate.

☑

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

Correct

Correct.

☑

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

Correct

Correct.