

1.

KNIME: In configuring the Numeric Binner node, what would happen if the definition for the humidity_low bin is changed from

] -infinity ... 25.0 [

to

] -infinity ... 25.0]

(i.e., the last bracket is changed from [to])?

☒ The definition for the humidity_low bin would change from excluding 25.0 to including 25.0

☐ The definition for the humidity_low bin would change from having 25.0 as the endpoint to having 25.1 as the endpoint

☐ Nothing would change

✓

Correct
2.

KNIME: Considering the Numeric Binner node again, what would happen if the “Append new column” box is not checked?

☒ The relative_humidity_3pm variable will become a categorical variable

☐ The relative_humidity_3pm variable will remain unchanged, and a new unnamed categorical variable will be created

☐ The relative_humidity_3pm variable will become undefined, and an error will occur

✓

Correct
3.

KNIME: How many samples had a missing value for air_temp_9am before missing values were addressed?

☒ 5

☐ 3

☐ 0

✓

Correct
4.

KNIME: How many samples were placed in the test set after the dataset was partitioned into training and test sets?

☒ 213

☐ 851

☐ 20

✓

Correct
5.

KNIME: What are the target and predicted class labels for the first sample in the test set?

☒ Both are humidity_not_low

☐ Target class label is humidity_not_low, and predicted class label is humidity_low

☐ Target class label is humidity_low, and predicted class label is humidity_not_low

✓

Correct
6.

Spark: What values are in the *number* column?

☒ Integer values starting at 0

☐ Time and date values

☐ Random integer values

✓

Correct
7.

Spark: With the original dataset split into 80% for training and 20% for test, how many of the first 20 samples from the test set were correctly classified?

☒ 16

☐ 10

☐ 1

✓

Correct
8.

Spark: If we split the data using 70% for training data and 30% for test data, how many samples would the training set have (using seed 13234)?

☒ 731

☐ 334

☐ 70

✓

Correct