1. Given an SQL table 'Books' with fields Title, Author, and Genre, which query would return a list of unique Genre values?
A.   SELECT Genre FROM Books;
B.   SELECT SET Genre FROM Books;
C.   SELECT UNIQUE Genre FROM Books;
D.   SELECT DISTINCT Genre FROM Books;

2. What are two reasons a data point would be treated as an outlier?
A.   If the value is greater than mean
B.   If the value is greater than median
C.   If the value is greater than standard deviation
D.   If the value is below the upper end of the bottom quartile by more then 1.5 times the interquartile range
E.   If the value is above the lower end of the top quartile by more then 1.5 times the interquartile range

3. What is the reason oversampling on the minority class should be performed after the train test is split and not before?
A.   To not risk reducing the value of the recall
B.   To not risk reducing the value of the accuracy
C.   To not increase the number of records in the testing dataset
D.   So there is not duplicate records in both the training and testing datasets

4. A Logistic Regression algorithm is used to classify images into four categories. If each image has a 5x5 pixel dimension, what is the the number of weights required (excluding biases) for this model?
A.   100
B.   200
C.   300
D.   400

5. K-nearest neighbors and K-means clustering with Euclidean distance suffer from the curse of dimensionality.

Is the above statement true and why?
A.   This is true. Euclidean distance, in general, is not a good metric in a

high-dimensional space.

B.   This is false. K-nearest neighbors can effectively work with a high-dimension space if K is small.

C.   This is true. As K increases, the number of samples required grows linearly with respect to the dimension and K.

D.   This is false. K-means clustering can effectively work with a high-dimension space if given a large number of samples.

6. What is one way that IBM AutoAI helps make it easier for data scientists to determine what key fields to use to join data tables?

A.   AutoAI automatically suggests joining on fields in different tables that have matching names.

B.   AutoAI automatically suggests joining on fields in different tables that have matching values.

C.   AutoAI automatically suggests joining on fields in different tables when those tables have the same number of rows.

D.   AutoAI automatically suggests joining on fields in different tables when those tables have the same number of fields.

7. The Bayesian optimization algorithm for hyperparameter tuning has which trait?

A.   It can be used for both regression and clustering but cannot be used for classification.

B.   It requires that a user provide a list of all of the values that will be tested for each hyperparameter.

C.   It builds a machine learning model to predict which hyperparameter settings will be most effective.

D.   If sufficient hardware is available, it can try out many different hyperparameter settings at the same time.

8. What does an R^2 Score (coefficient of determination) measure?

A.   It computes the average squared difference between estimated values and the actual value.

B.   It computes the square root of the average squared difference between estimated values and the actual value.

C.   It represents the proportion of the variance in the independent variable that has been explained by the dependent variable.

D.   It represents the proportion of the variance in the dependent variable that has been explained by the independent variable(s).

9. Converting a neural network into the newest version of TensorFlow or another deep-learning package is what type of performance drift or software decay?
A.   Data changes
B.   Concept drift
C.   Software changes
D.   Sampling bias and selection bias changes

See key below…

KEY:
1. D
2. D & E
3. D
4. A
5. A
6. A
7. C
8. D
9. C