Quiz 6 - Classification

1. Which of the following is a TRUE statement about classification?

- **Classification is a supervised task.**
- Classification is an unsupervised task.
- In a classification problem, the target variable has only two possible outcomes.

2. In which phase are model parameters adjusted?

- Testing phase
- **Training phase**
- Data preparation phase
- Model parameters are constant throughout the modeling process.

3. Which classification algorithm uses a probabilistic approach?

- **naive bayes**
- none of the above
- decision tree
- k-nearest-neighbors

4. What does the 'k' stand for in k-nearest-neighbors?

- the number of samples in the dataset
- **the number of nearest neighbors to consider in classifying a sample**
- the distance between neighbors: All neighboring samples that are 'k' distance apart from the sample are considered in classifying that sample.
- the number of training datasets

5. During construction of a decision tree, there are several criteria that can be used to determine when a node should no longer be split into subsets. Which one of the following is NOT applicable?

- The tree depth reaches a maximum threshold.
- The number of samples in the node reaches a minimum threshold.
- All (or X% of) samples have the same class label.
- **The value of the Gini index reaches a maximum threshold.**

6. Which statement is true of tree induction?

- You want to split the data in a node into subsets that are as homogeneous as possible
- **All of these statements are true of tree induction.**
- An impurity measure is used to determine the best split for a node.
- For each node, splits on all variables are tested to determine the best split for the node.

7. What does 'naive' mean in Naive Bayes?

- The full Bayes' Theorem is not used. The 'naive' in naive bayes specifies that a simplified version of Bayes' Theorem is used.
- The Bayes' Theorem makes estimating the probabilities easier. The 'naïve' in the name of classifier comes from this ease of probability calculation.
- **The model assumes that the input features are statistically independent of one another. The 'naïve' in the name of classifier comes from this naïve assumption.**

8. The feature independence assumption in Naive Bayes simplifies the classification problem by

- assuming that the prior probabilities of all classes are independent of one another.
- assuming that classes are independent of the input features.
- ignoring the prior probabilities altogether.
- **allowing the probability of each feature given the class to be estimated individually.**