✓ **Congratulations! You passed!**

Grade	Latest Submission	To pass 80% or	
received 100%	Grade 100%	higher	**Go to next item**

---

1. Which of the following statements accurately describe feature engineering? Select all that apply.

☐ Feature engineering does not involve using a data professional's statistical knowledge.

☑ In feature engineering, feature selection involves choosing the features in the data that contribute the most to predicting the response variable.

✓ **Correct**

☑ Feature engineering may involve transforming the properties of raw data.

✓ **Correct**

☑ In feature engineering, feature extraction involves taking multiple features to create a new one that will improve the accuracy of the algorithm.

✓ **Correct**

1 / 1 point

2. A data professional resolves a class imbalance in a very large dataset. They alter the majority class by using fewer of the original data points in order to produce a split that is more even. What does this scenario describe?

○ Merging

◉ Downsampling

○ Upsampling

○ Smoothing

✓ **Correct**

1 / 1 point

3. Fill in the blank: Customer _____ is the business term that describes how many customers stop using a product or service, or stop doing business with a company altogether, and at what rate this occurs.

◉ churn

○ retention

○ exchange

○ transfer

✓ **Correct**

1 / 1 point

4. Naïve Bayes's theorem enables data professionals to calculate posterior probability for a data project. What does posterior probability describe?

○ The likelihood of an event occurring based upon only observations and information that align with current hypotheses

○ The likelihood of an event occurring based upon the observations and information that were available at the start of the data project

◉ The likelihood of an event occurring after taking into consideration all new, relevant observations and information

○ The likelihood of an event occurring after taking into consideration only the most suitable observations and information

✓ **Correct**

1 / 1 point

5. Fill in the blank: When normalizing the columns in a dataset using MinMaxScaler, the columns' maximum value scales to one, and the minimum value scales to _____. Everything else falls somewhere in between.

◉ 0

○ -1

○ 0.1

○ .5

✓ **Correct**

1 / 1 point

6. A data professional explores a dataset by examining patterns to reveal key details about the data that will help inform the plans for building a model. Which PACE stage does this scenario describe?

○ Plan

○ Construct

○ Execute

◉ Analyze

✓ **Correct**

1 / 1 point

7. In the model-development process, which type of feature does not contain any useful information for predicting the target variable?

◉ Irrelevant

○ Predictive

○ Relevant

○ Conducive

✓ **Correct**

1 / 1 point

8. Fill in the blank: Log normalization is useful when working with a model that cannot manage continuous variables with _____ distributions.

○ normal

○ binomial

○ probability

◉ skewed

✓ **Correct**

1 / 1 point

9. A data professional is working on a cybersecurity project to detect malware on computers. The dataset contains information on 10,000 computer scans, but only 700 of them identified malware. This means there are few instances that have malware. What does this say about the data and the evaluation metrics?

1 / 1 point

○ There might be errors in the data, and you should clean it before using it.

◉ The data is imbalanced, and precision and recall metrics are useful for evaluating how well the model handles this issue.

○ The data is perfectly balanced and accuracy is the best metric to evaluate your model.

○ The data is too small, and you should collect more information before proceeding.

⊘ **Correct**

10. You are trying to find the probability of an event, A, given that another event, B is true. Which theorem could you use in this situation?

1 / 1 point

○ Central Limit Theorem

◉ Bayes Theorem

○ No Free Lunch Theorem

○ Law of Large Numbers

⊘ **Correct**

9. A data professional is working on a cybersecurity project to detect malware on computers. The dataset contains information on 10,000 computer scans, but only 700 of them identified malware. This means there are few instances that have malware. What does this say about the data and the evaluation metrics?

○ There might be errors in the data, and you should clean it before using it.

◉ The data is imbalanced, and precision and recall metrics are useful for evaluating how well the model handles this issue.

○ The data is perfectly balanced and accuracy is the best metric to evaluate your model.

○ The data is too small, and you should collect more information before proceeding.

⊘ **Correct**