

PRACTICE QUESTIONS

Pega Certified Data Scientist

SHARED BY [u/golden_gecko_59](#) on aiexamcerts

EDITION	Free Set 1
QUESTIONS	10
SHARED	October 15, 2026
EXAM	Pega Certified Data Scientist

Shared by [u/golden_gecko_59](#) on aiexamcerts

10 practice questions for the Pega Certified Data Scientist exam, each followed by its answer and a short explanation. More free sets and the full practice exam:

www.aiexamcerts.com/pega/pega-certified-data-scientist. Spotted a mistake? Email support@aiexamcerts.com with the exam name and the question number.

1. A marketing team wants the most valuable relevant offer for each customer, not only the most likely one. Which element weighs business value together with propensity?
- A Prioritization, which combines value, propensity and context weight
 - B An engagement policy only, which blocks offers
 - C Text categorization of the offer names, which sorts each offer into a product family for reporting
 - D Champion-challenger splitting only

Answer: A

Prioritization combines business value, propensity and context weight, so the highest-value relevant action can win even when its propensity is lower. An engagement policy is the most tempting distractor, since it controls which offers reach a customer, but it limits contact rather than ranking value. Text categorization and champion-challenger splitting serve other purposes.

2. AUC is the area under the ROC curve, which plots the true positive rate against the false positive rate across all thresholds. The full name of the curve is the receiver operating _____.

Answer: characteristic

The ROC curve is named the receiver operating characteristic, and AUC summarizes how well a model separates responders from non-responders across thresholds. Precision and recall curves are a related tool, but they are not the ROC curve that AUC describes.

3. A customer declined an offer last week. What should the engagement policy usually do?
- A Show the same offer again in every channel immediately, so the customer sees it wherever they go next
 - B Increase the offer's value to change the customer's mind
 - C Delete the customer record from the decision data
 - D Suppress the same offer for a defined period so the customer is not over-contacted

Answer: D

Engagement policies limit contact frequency and suppress repeated offers, which protects the relationship and the response rate over time. Immediate repetition is the most tempting distractor for teams that want to maximize contact, but it tends to lower responses. Deleting records and raising the value do not do what suppression does.

4. An adaptive model that never receives fresh responses will keep improving its predictions automatically over time.

A True
B False

Answer: B (False)

This is false. Adaptive models improve only as new outcomes arrive. Without fresh responses they stop learning, and their predictions can become stale as customer behavior changes.

5. Which two statements about training and evaluating predictive models are correct? (Choose two.)

A A holdout set estimates performance on data the model did not train on
B Evaluating on the training data gives an unbiased estimate of future performance
C Imbalanced outcomes can make overall accuracy misleading
D Removing the test data after training improves the model's accuracy
E Adding the outcome field as a predictor always improves the model

Answer: A, C

A holdout set estimates how the model performs on records it has not seen, and imbalanced outcomes can make accuracy look strong for the wrong reasons. Evaluating on training data is the most tempting distractor, because it gives a high score, but that score mostly reflects memorization. Removing test data and using the outcome as a predictor are both wrong, and the second is leakage.

6. A churn model uses the number of support calls in the last 30 days, but in production that data arrives a week late. What is the main risk?

A The feature may be unavailable or stale at prediction time, so training results overstate production performance
B The model will become more accurate because it sees fresher data
C The feature must be converted to text before scoring, which keeps the numeric values from being used by the model at all
D The churn label will be copied into the feature automatically

Answer: A

If a feature is not available when predictions are made, the training inputs differ from the production inputs and the offline score overstates real performance. Fresher data is the most tempting distractor, since it sounds like an improvement, but the problem here is the delay. Text conversion and label copying do not relate to the timing issue.

7. A call-center team wants to know whether each customer message is positive, negative or neutral. Which NLP technique fits?
- A Entity extraction of account numbers
 - B Sentiment analysis
 - C Propensity modeling
 - D Arbitration of conflicting actions

Answer: B

Sentiment analysis labels text by the opinion it expresses, which matches the positive, negative or neutral question. Entity extraction is the most tempting distractor, because it also reads customer messages, but it finds names, numbers and other items rather than tone. Propensity modeling and arbitration are decision-stage techniques.

8. Adaptive models in Pega learn from customer _____, such as accepted or rejected offers, as they occur.

Answer: responses (also accepted: outcomes, reactions)

Responses or outcomes recorded after an action are the signal that adaptive models use to update their predictions. Models without that feedback cannot adapt.

9. Which signal best detects data drift in the inputs of a deployed model?
- A The number of rows in the production table
 - B The average response time of the scoring service
 - C Changes in the distribution of input values compared with the training data
 - D The ratio of positive outcomes in the training sample

Answer: C

Data drift appears as a shift in the distribution of input values, which can be measured against the training baseline. Row count is the most tempting distractor, since volume changes are easy to see, but a bigger table can still have the same distribution. Scoring response time and the training outcome ratio do not measure a shift in the inputs.

10. In text analytics, what is the purpose of stop words?

- A The most important words, which are always kept in the model and weighted above every other term in the vocabulary
- B Common words such as "the" and "and", which are often removed because they carry little meaning
- C Words that are translated into a separate language before scoring
- D Words that are replaced by their synonyms before the text is vectorized

Answer: B

Stop words carry little meaning and are often removed so the informative terms stand out. Keeping the important words is the most tempting distractor, because common and important words can both appear near the top of a frequency list, but important words are the ones that are kept. Translation and synonym replacement have no role in this step.