

PRACTICE QUESTIONS

Pega Certified Data Scientist

SHARED BY [u/Mellow-Wombat-1378](#) on aiexamcerts

EDITION	Free Set 2
---------	------------

QUESTIONS	10
-----------	----

SHARED	October 15, 2026
--------	------------------

EXAM	Pega Certified Data Scientist
------	-------------------------------

Shared by [u/Mellow-Wombat-1378](#) 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. In a decision strategy, which component ranks the actions that remain after ineligible actions have been removed?
- A Categorize, which labels the customer's messages and routes each one to a handling queue
 - B Filter, which removes ineligible actions
 - C Archive, which stores each decision for later
 - D Prioritize, which ranks the actions that remain eligible

Answer: D

Prioritize ranks the eligible actions by their value and propensity, which determines the order in which they are offered. Filter is the most tempting distractor, since it runs just before ranking, but filtering only removes actions and does not order them. Archive and categorize serve other purposes.

2. Feature scaling is never needed for tree-based predictive models, because they split on value thresholds rather than on distances.
- A True
 - B False

Answer: A (True)

This is true. Tree-based models choose thresholds on each feature separately, so rescaling a feature does not change the splits. Distance-based and linear models usually do need scaling, which is why the choice of algorithm matters.

3. Which practice matters most when splitting time-based customer data for model evaluation?
- A Split by time so that the evaluation period follows the training period
 - B Split at random so that every month appears in both sets equally
 - C Use the same records for training and evaluation to keep results consistent
 - D Put the newest records in training and the oldest in evaluation

Answer: A

A time-based split mirrors production, where the model predicts future behavior from past data. Random splitting is the most tempting distractor for teams used to it, but it lets the model see the future. Reversing the order and reusing the same records both give unrealistic results.

4. The share of flagged cases that turn out to be truly positive is called _____.

Answer: precision

Precision measures the proportion of flagged cases that are correct, which matters when false alarms are costly. Recall is the related measure that counts how many real positives are found, so the two are often traded against each other.

5. Which two statements about monitoring a deployed model are correct? (Choose two.)

- A Monitoring should track both input drift and outcome performance
- B A model that passed review once needs no further monitoring
- C Performance should be compared by customer segment as well as overall
- D Monitoring is only needed for models that use text
- E Alerts can be ignored whenever overall accuracy looks stable

Answer: A, C

Input drift can appear before outcomes change, and segment views catch problems that averages hide. Skipping monitoring after one review is the most tempting distractor, because teams want to move on, but production data keeps changing. Text-only monitoring and ignoring alerts are both unsafe.

6. A champion strategy has been the default for months, and a new challenger shows a higher response rate in a small test. What should happen next?

- A Replace the champion immediately for all customers, since a higher response rate in any test proves the challenger is better
- B Delete the champion so there is no comparison
- C Keep testing the challenger on a defined audience and promote it only if the result holds
- D Stop measuring responses until the next quarter

Answer: C

A defined test with enough responses shows whether the lift is real before any promotion. Immediate replacement is the most tempting distractor, since a higher rate looks like a win, but a small sample can mislead. Deleting the champion removes the baseline, and pausing measurement hides the result.

7. Which Process AI use case is the best fit for a predictive model?
- A Sending the same reminder to every customer each day, which looks like automation even though it uses no prediction at all
 - B Storing a copy of every attachment as a backup
 - C Archiving resolved cases once the retention period has ended
 - D Predicting whether a service case is likely to breach its service level so it can be escalated early

Answer: D

A breach prediction gives the workflow a decision point, so the case can be escalated before the deadline is missed. Sending the same reminder to everyone is the most tempting distractor, because it looks like automation, but it does not use a prediction to choose an action. Backups and archiving are operational tasks rather than predictions.

8. Named entity extraction identifies items such as people, organizations and dates in text, and it is a different task from sentiment analysis.
- A True
 - B False

Answer: A (True)

This is true. Entity extraction finds and labels the named items in text, while sentiment analysis judges the tone of the text. The two are often used together but answer different questions.

9. Which step turns a text document into numeric features a model can use, such as term counts or TF-IDF weights?
- A Vectorization of the text into term counts or TF-IDF weights
 - B Arbitration of competing actions across every strategy in the decision hub during each customer interaction
 - C Eligibility filtering of customer records
 - D Scheduling a batch run every night

Answer: A

Vectorization converts words into numeric features, such as term counts or TF-IDF weights, that a model can learn from. Arbitration is the most tempting distractor for Pega readers, because it also resolves between options, but it works on decisions rather than on text. Eligibility filtering and scheduling are unrelated to text features.

10. Why do teams often wait for a minimum number of responses before trusting an adaptive model's results?
- A To make the model run faster on the server by skipping the outcome data needed for learning
 - B To avoid acting on unreliable early estimates from too little outcome data
 - C To reduce the number of retraining jobs in the schedule
 - D To remove the need for any engagement policy

Answer: B

Small samples produce unstable estimates, so waiting for enough responses reduces wrong decisions. Speed is the most tempting distractor, since a threshold sounds like a performance setting, but its purpose is reliability. Retraining job counts and engagement policy are unrelated to the threshold.