



CompTIA DataAI DY0-001 Practice Questions

Shared by Gilbert on 13-08-2026

For More Free Questions and Preparation Resources

Check the Links on Last Page



Question 1

Question Type: MultipleChoice

A data scientist is building a model to predict customer credit scores based on information collected from reporting agencies. The model needs to automatically adjust its parameters to adapt to recent changes in the information collected. Which of the following is the best model to use?

Options:

- A- Decision tree
- B- Random forest
- C- Linear discrimination analysis
- D- XGBoost

Answer:

D

Explanation:

XGBoost supports "warm-start" incremental training, continuing to refine the existing ensemble with new data, so it can automatically update its parameters as new agency information arrives. The other methods require full retraining to incorporate recent changes.

Question 2

Question Type: MultipleChoice

A data analyst is analyzing data and would like to build conceptual associations. Which option best is the best way to accomplish this task?

Options:

- A- n-grams
- B- NER
- C- TF-IDF
- D- POS

Answer:

A

Explanation:

n-grams capture contiguous sequences of words, revealing which terms co-occur and form meaningful multi-word concepts. By analyzing these frequent word combinations, you directly uncover conceptual associations in the text.

Question 3

Question Type: MultipleChoice

A data scientist wants to evaluate the performance of various nonlinear models. Which of the following is best suited for this task?

Options:

- A- AIC
- B- Chi-squared test
- C- MCC
- D- ANOVA

Answer:

A

Question 4

Question Type: MultipleChoice

A data scientist wants to predict a person's travel destination. The options are:

Which of the following models would best fit this use case?

Options:

- A- Linear discriminant analysis
- B- k-means modeling
- C- Latent semantic analysis
- D- Principal component analysis

Answer:

A

Explanation:

You need a supervised multiclass classification model to predict one of the four labeled destinations. Linear Discriminant Analysis is designed for such tasks, finding the linear boundaries that best separate the known destination classes.

Question 5

Question Type: MultipleChoice

A data analyst wants to find the latitude and longitude of a mailing address. Which option best is the best method to use?

Options:

- A- One-hot encoding
- B- Binning
- C- Geocoding
- D- Imputing

Answer:

C

Explanation:

Geocoding is the process of converting a postal address into geographic coordinates (latitude and longitude), making it the appropriate method.

Question 6

Question Type: MultipleChoice

A data scientist is building an inferential model with a single predictor variable. A scatter plot of the independent variable against the real-number dependent variable shows a strong relationship between them. The predictor variable is normally distributed with very few outliers. Which of the following algorithms is the best fit for this model, given the data scientist wants the model to be easily interpreted?

Options:

- A- A logistic regression
- B- An exponential regression
- C- A linear regression
- D- A probit regression

Answer:

C

Question 7

Question Type: MultipleChoice

A data scientist is performing a linear regression and wants to construct a model that explains the most variation in the data

a. Which of the following should the data scientist maximize when evaluating the regression performance metrics?

Options:

- A- Accuracy
- B- R²
- C- p value
- D- AUC

Answer:

B

To Get Premium Files for DY0-001 Visit

<https://www.p2pexams.com/products/dy0-001>

For More Free Questions Visit

<https://www.p2pexams.com/comptia/pdf/dy0-001>

20%
DISCOUNT

P2P
exams