Data Mining Practice Test (Sample)

Study Guide

Everything you need from our exam experts!

Sample study guide. For the full version with hundreds of questions, visit: https://datamining.examzify.com

Copyright © 2026 by Examzify - A Kaluba Technologies Inc. product.

ALL RIGHTS RESERVED.

No part of this book may be reproduced or transferred in any form or by any means, graphic, electronic, or mechanical, including photocopying, recording, web distribution, taping, or by any information storage retrieval system, without the written permission of the author.

Notice: Examzify makes every reasonable effort to obtain accurate, complete, and timely information about this product from reliable sources.

SAMPLE

Table of Contents

SAMPLE

Introduction

Preparing for a certification exam can feel overwhelming, but with the right tools, it becomes an opportunity to build confidence, sharpen your skills, and move one step closer to your goals. At Examzify, we believe that effective exam preparation isn't just about memorization, it's about understanding the material, identifying knowledge gaps, and building the test-taking strategies that lead to success.

This guide was designed to help you do exactly that.

Whether you're preparing for a licensing exam, professional certification, or entry-level qualification, this book offers structured practice to reinforce key concepts. You'll find a wide range of multiple-choice questions, each followed by clear explanations to help you understand not just the right answer, but why it's correct.

The content in this guide is based on real-world exam objectives and aligned with the types of questions and topics commonly found on official tests. It's ideal for learners who want to:

- **Practice answering questions under realistic conditions,**
- **Improve accuracy and speed,**
- **Review explanations to strengthen weak areas, and**
- **Approach the exam with greater confidence.**

We recommend using this book not as a stand-alone study tool, but alongside other resources like flashcards, textbooks, or hands-on training. For best results, we recommend working through each question, reflecting on the explanation provided, and revisiting the topics that challenge you most.

Remember: successful test preparation isn't about getting every question right the first time, it's about learning from your mistakes and improving over time. Stay focused, trust the process, and know that every page you turn brings you closer to success.

Let's begin.

SAMPLE

How to Use This Guide

This guide is designed to help you study more effectively and approach your exam with confidence. Whether you're reviewing for the first time or doing a final refresh, here’s how to get the most out of your Examzify study guide:

1. Start with a Diagnostic Review

Skim through the questions to get a sense of what you know and what you need to focus on. Your goal is to identify knowledge gaps early.

2. Study in Short, Focused Sessions

Break your study time into manageable blocks (e.g. 30 – 45 minutes). Review a handful of questions, reflect on the explanations.

3. Learn from the Explanations

After answering a question, always read the explanation, even if you got it right. It reinforces key points, corrects misunderstandings, and teaches subtle distinctions between similar answers.

4. Track Your Progress

Use bookmarks or notes (if reading digitally) to mark difficult questions. Revisit these regularly and track improvements over time.

5. Simulate the Real Exam

Once you're comfortable, try taking a full set of questions without pausing. Set a timer and simulate test-day conditions to build confidence and time management skills.

6. Repeat and Review

Don’t just study once, repetition builds retention. Re-attempt questions after a few days and revisit explanations to reinforce learning. Pair this guide with other Examzify tools like flashcards, and digital practice tests to strengthen your preparation across formats.

There’s no single right way to study, but consistent, thoughtful effort always wins. Use this guide flexibly, adapt the tips above to fit your pace and learning style. You've got this!

SAMPLE

Questions

SAMPLE

1. **True or false: In a decision tree, some independent variables can be used more than once.**

   A. False

   B. True

   C. Sometimes

   D. Never

2. **Correlation is a statistical measure of how strong the relationships are between _____ in a data set.**

   A. Records

   B. Attributes

   C. Observations

   D. Clusters

3. **Data arranged into columns and rows in a database are stored in which structure?**

   A. Databases

   B. Tables

   C. Schemas

   D. Records

4. **True or false: The neural network is a classification model.**

   A. Not specified

   B. Sometimes

   C. True

   D. False

5. **Removing columns from a data set because they are not useful for a certain type of data analysis is an example of _____.**

   A. Data cleaning

   B. Data transformation

   C. Normalization

   D. Attribute reduction

SAMPLE

6. In order to describe data, we often use arithmetic functions such as the mean, median, and mode. These three functions are known as _____.

A. Measures of dispersion

B. Measures of shape

C. Measures of outliers

D. Measures of central tendency

7. True or false: The neural network is both a predictive data mining model and a classification model.

A. True

B. False

C. Not specified

D. Sometimes

8. Which term is NOT commonly used to refer to a single row in a data table?

A. Datum

B. Record

C. Row

D. Tuple

9. Which data type is explicitly required for independent variables in neural network modeling?

A. Categorical

B. Text

C. Binary

D. Numeric

10. Which values can be used for a binominal variable?

A. 0 and 1 only

B. All of the above

C. Only two values

D. Non-numeric values allowed

SAMPLE

Answers

SAMPLE

1. B
2. B
3. B
4. B
5. D
6. D
7. A
8. A
9. D
10. B

SAMPLE

Explanations

SAMPLE

1. True or false: In a decision tree, some independent variables can be used more than once.

A. False

B. True

C. Sometimes

D. Never

In a decision tree, using the same feature more than once is allowed and common. The tree splits data to increase purity, and there's no rule preventing a feature from appearing again in deeper nodes or in different branches. For numeric features, different thresholds can be chosen in different parts of the tree, refining the partition within specific subsets. For example, age might be split at 30 at the root, and within the younger subset you could split again on age at 18 to separate younger children from teens, or within the older subset split at 50 to separate middle-aged from older adults. This reuse helps the tree capture more nuanced patterns across different value ranges of the same variable. So the statement is true.

2. Correlation is a statistical measure of how strong the relationships are between _____ in a data set.

A. Records

B. Attributes

C. Observations

D. Clusters

Correlation measures how strongly two attributes in a data set vary together across all observations. It captures both the strength and the direction of their linear relationship, so it's about relationships between variables (columns) rather than about individual data points (rows) or about grouping data into clusters. This is why attributes is the best fit. For example, height and weight tend to rise together, showing a positive correlation, while two unrelated attributes would show little or no correlation.

3. Data arranged into columns and rows in a database are stored in which structure?

A. Databases

B. Tables

C. Schemas

D. Records

In relational databases, the grid-like arrangement of data with columns representing attributes and rows representing records is stored in a table. A table provides the structure for a set of similar records: each column defines a data type and attribute, while each row holds a complete instance of those attributes. This is why data organized by columns and rows is described as being stored in a table. It's useful to keep in mind related terms: a database is the collection of tables (and other objects); a schema defines the structure and relationships of those tables; a record (or row) is a single entry within a table.

SAMPLE

4. True or false: The neural network is a classification model.

A. Not specified

B. Sometimes

C. True

D. False

Neural networks are versatile models that can be used for many tasks, not just one. They can function as classifiers when their outputs represent class probabilities and they're trained with a classification loss like cross-entropy. They can also function as regressors when the outputs are continuous values and trained with a regression loss like mean squared error. Because of this flexibility, calling a neural network strictly a classification model isn't accurate; it's sometimes used for classification, but not exclusively. For example, a network designed to recognize digits uses a softmax output and cross-entropy loss, while a network predicting house prices uses a linear output with a regression loss. So the statement is true only in some contexts, making "Sometimes" the best answer.

5. Removing columns from a data set because they are not useful for a certain type of data analysis is an example of _____.

A. Data cleaning

B. Data transformation

C. Normalization

D. Attribute reduction

Removing columns that aren't useful for a specific analysis targets reducing the number of attributes we consider, which is known as attribute reduction. This is a form of feature selection: we keep only the informative attributes and drop the rest. By pruning irrelevant or redundant columns, the model becomes simpler, runs faster, and often generalizes better because it's less affected by noise. This approach is different from data cleaning (fixing errors or missing values), data transformation (changing the representation of data, like encoding or scaling), and normalization (scaling features to a common range). The core idea here is to streamline the feature set by removing unnecessary columns to focus on the information that matters for the analysis.

SAMPLE

6. In order to describe data, we often use arithmetic functions such as the mean, median, and mode. These three functions are known as _____.

A. Measures of dispersion

B. Measures of shape

C. Measures of outliers

D. Measures of central tendency

These statistics describe where data tend to cluster, capturing a single representative value for the data set. The mean, median, and mode are all ways to describe the central location of the distribution, so they are called measures of central tendency. The mean is the arithmetic average, the median is the middle value when the data are ordered, and the mode is the most frequent value. They each summarize the data's center in a slightly different way, which is why all three are grouped together as central tendency. It helps to contrast this with measures of dispersion (how spread out the data are, like range or standard deviation) or measures of shape (the distribution's symmetry or peakedness). For example, in a skewed distribution, the mean can be pulled toward the tail, while the median remains closer to the bulk of the data, illustrating how they describe center from different perspectives.

7. True or false: The neural network is both a predictive data mining model and a classification model.

A. True

B. False

C. Not specified

D. Sometimes

Neural networks are flexible function approximators used in data mining for both predicting numeric values and assigning instances to categories. They learn a mapping from input features to outputs by adjusting weights to minimize error on training data, which makes them a predictive model in general. For regression tasks, the network outputs a continuous value, often with a linear or identity activation at the final layer. For classification tasks, the final layer uses a probability-friendly activation like softmax (for multiple classes) or sigmoid (for binary), turning the network into a classifier. Because the same architecture can be trained for either objective by choosing the appropriate loss function and outputs, neural networks can serve as both predictive models and classification models.

SAMPLE

8. Which term is NOT commonly used to refer to a single row in a data table?

A. Datum

B. Record

C. Row

D. Tuple

In data tables, a single row represents one entity's entire set of values across all attributes, so the common terms for that unit are a row, a record, or a tuple. A datum, on the other hand, is a single value or data item within the row, not the whole row itself. So the term that is not used to refer to the entire row is datum, since it denotes an individual piece of data rather than the complete collection of values for one observation. The other terms fit because they convey the idea of the whole row of values for one entity.

9. Which data type is explicitly required for independent variables in neural network modeling?

A. Categorical

B. Text

C. Binary

D. Numeric

Neural networks operate on numbers. The computations inside a network—multiplying inputs by weights, summing them, applying activation functions, and backpropagating errors—rely on numeric values and differentiable math. Because of that, the independent variables you feed into a network must be numeric data types. If you start with non-numeric data, you convert it into numbers first. Categorical features are encoded (for example, one-hot or label encoding) so they become numeric vectors. Text data is transformed into numeric representations through methods like embeddings or vectorization. Binary data is already numeric (0 or 1), so it fits naturally. In short, the explicit requirement is numeric input, with non-numeric data needing conversion before modeling.

10. Which values can be used for a binominal variable?

A. 0 and 1 only

B. All of the above

C. Only two values

D. Non-numeric values allowed

Binary (binomial) variables represent two outcomes and are kept numeric so we can perform math and probability calculations. The standard coding is 0 for one outcome and 1 for the other, so the variable can take exactly two numeric values. That's why it's correct to say it can be 0 and 1, or simply that it has two values. Non-numeric values wouldn't work in this context, because you need numbers to compute things like sums, averages, and probabilities. Therefore the statements describing two numeric values, typically 0 and 1, reflect the correct concept. An option claiming all of the above would be incorrect due to the non-numeric possibility.

SAMPLE

Next Steps

Congratulations on reaching the final section of this guide. You've taken a meaningful step toward passing your certification exam and advancing your career.

As you continue preparing, remember that consistent practice, review, and self-reflection are key to success. Make time to revisit difficult topics, simulate exam conditions, and track your progress along the way.

If you need help, have suggestions, or want to share feedback, we'd love to hear from you. Reach out to our team at hello@examzify.com.

Or visit your dedicated course page for more study tools and resources:

https://datamining.examzify.com

We wish you the very best on your exam journey. You've got this!

SAMPLE