DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST^Q&As

Databricks Certified Professional Data Scientist Exam

Pass Databricks DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Exam with 100% Guarantee

Free Download Real Questions & Answers **PDF** and **VCE** file from:

https://www.passapply.com/databricks-certified-professional-data-scientist.html

100% Passing Guarantee
100% Money Back Assurance

Following Questions and Answers are all new published by Databricks Official Exam Center

DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST VCE Dumps | DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-SCIENTIST Exam Questions |
DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Braindumps

2 / 5

QUESTION 1

Which of the following metrics are useful in measuring the accuracy and quality of a recommender system?

A. Cluster Density

B. Support Vector Count

C. Mean Absolute Error

D. Sum of Absolute Errors

Correct Answer: C

Explanation: The MAE measures the average magnitude of the errors in a set of forecasts, without considering their direction. It measures accuracy for continuous variables. The equation is given in the library references. Expressed in words,

the MAE is the average over the verification sample of the absolute values of the differences between forecast and the corresponding observation. The MAE is a linear score which means that all the individual differences are weighted equally

in the average. The sum of absolute errors is a valid metric, but doesn\\'t give any useful sense of how the recommender system is performing.

Support vector count and cluster density do not apply to recommender systems. MAE and AUC are both valid and useful metrics for measuring recommender systems.

QUESTION 2

In which phase of the data analytics lifecycle do Data Scientists spend the most time in a project?

A. Discovery

B. Data Preparation

C. Model Building

D. Communicate Results

Correct Answer: B

QUESTION 3

You have collected the 100\\'s of parameters about the 1000\\'s of websites e.g. daily hits, average time on the websites, number of unique visitors, number of returning visitors etc. Now you have find the most important parameters which can best describe a website, so which of the following technique you will use:

A. PCA (Principal component analysis)

B. Linear Regression

[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST VCE Dumps](link) | [DATABRICKS-CERTIFIED](link)
[PROFESSIONAL-DATA-SCIENTIST Exam Questions](link) |
[DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Braindumps](link)

3 / 5

C. Logistic Regression

D. Clustering

Correct Answer: A

Explanation: Principal component analysis . or PCA, is a technique for taking a dataset that is in the form of a set of tuples representing points in a high-dimensional space and finding the dimensions along which the tuples line up best. The idea is to treat the set of tuples as a matrix M and find the eigenvectors for MMT or M T M . The matrix of these eigenvectors can be thought of as a rigid rotation in a high-dimensional space. When you apply this transformation to the original data, the axis corresponding to the principal eigenvector is the one along which the points are most "spread out,11 More precisely this axis is the one along which the variance of the data is maximized. Put another way, the points can best be viewed as lying along this axis, with small deviations from this axis.

QUESTION 4

What describes a true property of Logistic Regression method?

A. It handles missing values well.

B. It works well with discrete variables that have many distinct values.

C. It is robust with redundant variables and correlated variables.

D. It works well with variables that affect the outcome in a discontinuous way.

Correct Answer: C

QUESTION 5

What are the advantages of the Hashing Features?

A. Requires the less memory

B. Less pass through the training data

C. Easily reverse engineer vectors to determine which original feature mapped to a vector location

Correct Answer: AB

Explanation: SGD-based classifiers avoid the need to predetermine vector size by simply picking a reasonable size and shoehorning the training data into vectors of that size. This approach is known as feature hashing. The shoehorning is

done by picking one or more locations by using a hash of the name of the variable for continuous variables or a hash of the variable name and the category name or word for categorical, text-like, or word-like data.

This hashed feature approach has the distinct advantage of requiring less memory and one less pass through the training data, but it can make it much harder to reverse engineer vectors to determine which original feature mapped to a vector

location. This is because multiple features may hash to the same location. With large vectors or with multiple locations per feature, this isn\\'t a problem for accuracy but it can make it hard to understand what a classifier is doing.

An additional benefit of feature hashing is that the unknown and unbounded vocabularies typical of word-like variables aren\\'t a problem.

<div>

DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-
SCIENTIST VCE Dumps

DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-
SCIENTIST Exam
Questions

DATABRICKS-CERTIFIED-
PROFESSIONAL-DATA-
SCIENTIST Braindumps

</div>