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.leads4pass.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

QUESTION 1

You are working on a problem where you have to predict whether the claim is done valid or not. And you find that most of the claims which are having spelling errors as well as corrections in the manually filled claim forms compare to the honest claims. Which of the following technique is suitable to find out whether the claim is valid or not?

A. Naive Bayes

B. Logistic Regression

C. Random Decision Forests

D. Any one of the above

Correct Answer: D

Explanation: In this problem you have been given high-dimensional independent variables like texts, corrections, test results etc. and you have to predict either valid or not valid (One of two). So all of the below technique can be applied to this problem. Support vector machines Naive Bayes Logistic regression Random decision forests

---

QUESTION 2

What is one modeling or descriptive statistical function in MADlib that is typically not provided in a standard relational database?

A. Expected value

B. Variance

C. Linear regression

D. Quantiles

Correct Answer: C

Explanation: Linear regression models a linear relationship of a scalar dependent variable y to one or more explanatory independent variables x to build a model of coefficients.

QUESTION 3

Which of the below best describe the Principal component analysis

A. Dimensionality reduction

B. Collaborative filtering

C. Classification

D. Regression

E. Clustering

Correct Answer: A

QUESTION 4

In which of the scenario you can use the linear regression model?

A. Predicting Home Price based on the location and house area

B. Predicting demand of the goods and services based on the weather

C. Predicting tumor size reduction based on input as number of radiation treatment

D. Predicting sales of the text book based on the number of students in state

Correct Answer: ABCD

Explanation: : You can use the linear regression model for predicting the continuous output variable based on the input variables. In all the cases mentioned in the question option, you can see that output can be predicted based on the input variable. Option-A: Input: Location, House Area and Output: House Price Option-B : Input: Weather condition, Output: Demand for the goods and services Option-C : Input: Number of Radiation Session Output: Tumor Size Reduction Option-D : Input: Number of students and Output: Sale quantity of text book

QUESTION 5

In which lifecycle stage are test and training data sets created?

A. Model planning

B. Discovery

C. Model building

D. Data preparation

Correct Answer: C

Explanation: In Phase 1, the team learns the business domain, including relevant history such as whether the organization or business unit has attempted similar projects in the past from which they can learn. The team assesses the resources available to support the project in terms of people, technology time, and data. Important activities in this phase include framing the business problem as an analytics challenge that can be addressed in subsequent phases and formulating initial hypotheses (IHs) to test and begin learning the data. Data preparation: Phase 2 requires the presence of an analytic sandbox, in which the team can work with data and perform analytics for the duration of the project. The team needs to execute extract, load, and transform (ELT) or extract, transform and load (ETL) to get data into the sandbox. The ELT and ETL are sometimes abbreviated as ETLT Data should be transformed in the ETLT process so the team can work with it and analyze it. In this phase, the team also needs to familiarize itself with the data thoroughly and take steps to condition the data Model planning: Phase 3 is model planning, where the team determines the methods, techniques, and workflow it intends to follow for the subsequent model building phase. The team explores the data to learn about the relationships between variables and subsequently selects key variables and the most suitable models. Model building: In Phase 4, the team develops datasets for testing, training, and production purposes. In addition, in this phase the team builds and executes models based on the work done in the model planning phase. The team also considers whether its existing tools will suffice for running the models, or if it will need a more robust environment for executing models and workflows (for example, fast hardware and parallel processing, if applicable). Communicate results: In Phase 5, the team, in collaboration with major stakeholders, determines if the results of the

project are a success or a failure based on the criteria developed in Phase 1. The team should identify key findings, quantify the business value, and develop a narrative to summarize and convey findings to stakeholders. Operationalize: In Phase 6, the team delivers final reports, briefings, code, and technical documents. In addition, the team may run a pilot project to implement the models in a production environment.