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.pass2lead.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

Find out the classifier which assumes independence among all its features?

A. Neural networks

B. Linear Regression

C. Naive Bayes

D. Random forests

Correct Answer: C

Explanation: A Bayes classifier is a simple probabilistic classifier based on applying Bayes\\' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". A Bayes classifier is a simple probabilistic classifier based on applying Bayes\\' theorem (from Bayesian statistics) with strong (naive) independence assumptions. A more descriptive term for the underlying probability model would be "independent feature model". In simple terms, a naive Bayes classifier assumes that the presence (or absence) of a particular feature of a class is unrelated to the presence (or absence) of any other feature. For example, a fruit may be considered to be an apple if it is red, round, and about 4" in diameter Even if these features depend on each other or upon the existence of the other features, a naive Bayes classifier considers all of these properties to independently contribute to the probability that this fruit is an apple.

QUESTION 2

Of all the smokers in a particular district, 40% prefer brand A and 60% prefer brand B. Of those smokers who prefer brand A. 30% are females, and of those who prefer brand B. 40% are female. What is the probability that a randomly selected smoker prefers brand A, given that the person selected is a female?

Which of the following is a best way to solve this problem?

A. Bays Theorem

B. Poisson Distribution

C. Binomial Distribution

D. None of the above

Correct Answer: A

QUESTION 3

What are the advantages of the mutual information over the Pearson correlation for text classification problems?

A. The mutual information has a meaningful test for statistical significance.

B. The mutual information can signal non-linear relationships between the dependent and independent variables.

C. The mutual information is easier to parallelize.

D. The mutual information doesn\\'t assume that the variables are normally distributed.

Correct Answer: C

Explanation: A linear scaling of the input variables (that may be caused by a change of units for the measurements) is sufficient to modify the PCA results. Feature selection methods that are sufficient for simple distributions of the patterns belonging to different classes can fail in classification tasks with complex decision boundaries. In addition, methods based on a linear dependence (like the correlation) cannot take care of arbitrary relations between the pattern coordinates and the different classes. On the contrary, the mutual information can measure arbitrary relations between variables and it does not depend on transformations acting on the different variables. This item concerns itself with feature selection for a text classification problem and references mutual information criteria. Mutual information is a bit more sophisticated than just selecting based on the simple correlation of two numbers because it can detect non- linear relationships that will not be identified by the correlation. Whenever possible: mutual information is a better feature selection technique than correlation. Mutual information is a quantification of the dependency between random variables. It is sometimes contrasted with linear correlation since mutual information captures nonlinear dependence. Correlation analysis provides a quantitative means of measuring the strength of a linear relationship between two vectors of data. Mutual information is essentially the measure of how much "knowledge" one can gain of a certain variable by knowing the value of another variable.

QUESTION 4

Suppose you have been given a relatively high-dimension set of independent variables and you are asked to come up with a model that predicts one of Two possible outcomes like "YES" or "NO", then which of the following technique best fit?

A. Support vector machines

B. Naive Bayes

C. Logistic regression

D. Random decision forests

E. All of the above

Correct Answer: E

Explanation: In this problem you have been given high-dimensional independent variables like yeS; nO; no English words , 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 5

Which activity is performed in the Operationalize phase of the Data Analytics Lifecycle?

A. Define the process to maintain the model

B. Try different analytical techniques

C. Try different variables

D. Transform existing variables

Correct Answer: A

Explanation: Operationalize In the final phase, the team communicates the benefits of the project more broadly and sets up a pilot project to deploy the work in a controlled way before broadening the work to a full enterprise or ecosystem of users. In Phase 4. the team scored the model in the analytics sandbox.