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

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF Dumps | DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide

1 / 5

QUESTION 1

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 2

You are working on a Data Science project and during the project you have been gibe a responsibility to interview all the stakeholders in the project. In which phase of the project you are?

A. Discovery

B. Data Preparations

C. Creating Models

D. Executing Models

E. Creating visuals from the outcome

F. Operationnalise the models

Correct Answer: A

Explanation: During the discovery phase you will be interviewing all the project stakeholders because they would be having quite a good amount of knowledge for the problem domain you will be working and you also interviewing project sponsors you will get to know what all are the expectations once project get completed. Hence, you will be noting down all the expectations from the project as well as you will be using their expertise in the domain.

QUESTION 3

A researcher is interested in how variables, such as GRE (Graduate Record Exam scores), GPA (grade point average) and prestige of the undergraduate institution, effect admission into graduate school. The response variable, admit/don\\'t admit, is a binary variable.

Above is an example of:

A. Linear Regression

B. Logistic Regression

C. Recommendation system

D. Maximum likelihood estimation

E. Hierarchical linear models

Correct Answer: B

Explanation: Logistic regression Pros: Computationally inexpensive, easy to implement, knowledge representation easy to interpret Cons: Prone to underfitting, may have low accuracy Works with: Numeric values, nominal values

QUESTION 4

A problem statement is given as below

Hospital records show that of patients suffering from a certain disease, 75% die of it. What is the probability that of 6 randomly selected patients, 4 will recover?

Which of the following model will you use to solve it.

A. Binomial

B. Poisson

C. Normal

D. Any of the above

Correct Answer: A

QUESTION 5

Select the correct statement regarding the naive Bayes classification:

A. it only requires a small amount of training data to estimate the parameters

B. Independent variables can be assumed

C. only the variances of the variables for each class need to be determined

![Pass2Lead](https://Pass2Lead.com)
D. for each class entire covariance matrix need to be determined

Correct Answer: ABC

Explanation: An advantage of naive Bayes is that it only requires a small amount of training data to estimate the parameters (means and variances of the variables) necessary for classification. Because independent variables are assumed, only the variances of the variables for each class need to be determined and not the entire covariance matrix.

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps

DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF Dumps

DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide

Latest DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Dumps | DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST PDF Dumps | DATABRICKS-CERTIFIED-PROFESSIONAL-DATA-SCIENTIST Study Guide

5 / 5