DMW BodhiTree MCQs

Value of pearson coefficient ranges between-

Answer

-1 to +1

What if pearson coefficient value is between marks and sports attributes is 0.9?

Answer

Positively Related

What can be the value of pearson's coefficient? Select ALL correct options.

Answer

"+1", "0", "-0.6"

If I debit my account with Rs. 5000/- then which of the bank's database is updated?

Answer

"Bank's Operational Database"

Which of the following are the subjects?

Answer

"Student", "Teacher", "Lab"

Which of the following is nonvolatile?

Answer

Data warehouse

Write operation can be performed in OLAP

Answer

False

Following are the properties of OLAP

Answer

“Used by Knowledge workers”, “Complex Query”

6 dimensional data can be viewed as series –

Answer

5-D cubes

Data Cube allows data to be viewed in-

Answer

Multiple dimensions

Highest level of Summarization called-?

Answer

Base

Which data model is used for OLTP? Select ALL correct options.

Answer

“Relational”, “ER”, “Network”

Star Schema has a problem of -

Answer

Redundancy

Which Schema has multiple dimension tables

Answer

“Star”, “Snoflake”, “Fact Constellation”

Which schema has multiple fact tables

Answer

“Fact Cons.”

Which Schema is suitable for large enterprise?

Answer

Fact Cons.

Which Schema is suitable for data mart?? Select ALL correct options.

Answer

“Star”, “Snowflake”

Identify the correct option about data warehouse data.

Answer

“Data Warehouse contains detail as well as summarized data.”

What type of attributes help to store aggregate data in data cube?

Answer

“Measurable attributes in fact table”

Data Cube is the Graphical representation of the ___

Answer

“Fact Table”

In the multidimensional model, the records areorganized into various dimensions, and each dimension includes multiple levels of abstraction described by ____

Answer

“Concept Hierarchy”

Identify the correct order of concept Hierarchy

Answer

“street < city < province_or_state< country.”

How many types of Data warehouse Architecture

Answer

“3”

What is the purpose of Staging Area in DWA ?

Answer

“To Process raw data before storing in warehouse”

How many maximum Components can be their in data warehouse Architecture ?

Answer

“5”

Which are the components of bottom tier of DWH multitier architecture ?

Answer

“Metadata repositery, datawarehouse server and data marts”

What is advantage of ROLAP ?

Answer

“Scalabilty”

Benifits of MOLAP ?

Answer

“Faster Computation”

What are components of middle tier of DWH multitier architecture?

Answer

OLAP servers

Types of OLAP servers?

Answer

“ROLAP”, “MOLAP”, “HOLAP”

Which is 5th Step of data warehouse design Process?

Answer

“Online Analytic Processing (OLAP) Cube”

Primary physical environments are—

Answer

“Development, testing, and production”

OLAP cubes help to organize

Answer

“Data in a multi-dimensional format”

Requirements Gathering stage should focus on?

Answer

Aligning department goals with the overall project

Three most popular data models for warehouses are:

Answer

“Snowflake Schema”, “Star Schema”, “Galaxy Schema”

Similarity values always range from [0, 1]

Answer

“TRUE”

Fill the correct option about symmetry property of similarity. s(p, q) = ? for all p and q.

Answer

“s(q, p)”

Select correct example of symmetric binary attribute

Answer

Gender: Male, Female

Select correct example of Asymmetric binary attribute

Answer

“Student Result: Pass, Fail”

Identify the correct example of Nominal Attributes.

Answer

“Income categories - HIGH, MEDIUM, LOW”

Consider the two objects i and j with nominal attributes, the dissimilarity between these objects are calculated using below equation: d(i,j)= (p-m)/p. In this formula what p and m represents?

Answer

m is the number of matches p is the total number of variables/features

When objects are represented using single attribute, the proximity value 1 indicates :

Answer

“Objects are similar”

Identify the name of the table used for measuring similarity between objects represenred using 2 or more binary attributes.

Answer

“Contegency Table”

Gender is the example of Asymmetric Binary Attribute.

Answer

“FALSE”

Identity correct equation of Jacard Coefficient

Answer

“J= f11/f01+f10+f11”

In the equation of Minskowski Distance , what is the set of values of r parameter?

Answer

“[1,2, infinity]”

What equation we get when r parameter =2 in Minskowski Distance formula.

Answer

“Euclidean distance”

Find the Manhattan distance between the given objects represented using single attribute. Employee1(56) Employee2(60)

Answer

“4”

Distance matrix is square matrix

Answer

“FALSE”

Find the Euclidean distance between objects X and Y. X= (6,4) Y= (2,7)

Answer

“5”

Transactions containing x and y / total transactions?

Answer

support formula

Transactions containing x and y / transactions contain x?

Answer

confidence formula

What is differential market basket analysis?

Answer

Comparision will results between different stores, between customers

Maximal Itemset is -

Answer

“An itemset is maximal frequent if none of its supersets are frequent.”

Closed Itemset

Answer

“An itemset is closed if none of its immediate supersets have same support count same as Itemset”

Apriori Algorithm use -

Answer

“Iterative approach”

Minimum support is 75% for 4 transaction database means values is

Answer

“3”

Apriori algorithm final output will be

Answer

“Association rules”

Collabrative filtering used for

Answer

“Recommendation”

Lift is

Answer

“Increase in sell of one item if will sell other item”

Limitation of Apriori

Answer

“Algorithm scans the database too many times”

Answer Description

Algorithm scans all items only once

Which of the following is direct application of frequent itemset mining??

Answer

Market Basket Analysis

FP growth is

Answer

“Frequent Pattern Growth Algorithm”

A Frequent Pattern set is built which will contain

Answer

“All the elements whose frequency is greater than or equal to the minimum support”

Root node for FP tree is always

Answer

“null”

What does FP growth algorithm do??

Answer

It mines all frequent patterns by constructing a FP tree

Which of the following is true?

Answer

Both apriori and FP-Growth uses horizontal data format

Rules at abstract level are more stronger than rules at specific level

Answer

Yes

In reduced support multi level association rules category support at more abstract level is reduced than more specific levels

Answer

No

Supervised learning takes input as unlableled data

Answer

False

Regression is kind of Supervised Learning

Answer

Yes

Decision Tree is a kind of -

Answer

Supervised Learning

In Decision Tree, Attributes can be at leaf nodes

Answer

No

Which of the following is not the stopping criteria in DT induction? All samples for a given node belong to the same class There are no remaining attributes for further partitioning There are no samples left

Answer

All above

KNN is called Lazy Learner Because_____

Answer

“Solution is provided at the query time”

KNN algorithm works with numerical data

Answer

“TRUE”

KNN is example of ___ technique

Answer

“Supervised”

Associative classifier is combination of __ and ___

Answer

“Association Rule mining , Classification”

Association Rule mining , Selects Rules with right side is ___

Answer

“Class Label”

Case Based Reasoning is called Lazy Learner Because_____

Answer

“little work is done offline, and all of the work is performed at query time.”

According to Aamodt and Plaza , the CBR working cycle comprising of which four R’s :

Answer

“RETRIEVE REUSE REVISE RETAIN”

Entropy Value ranges between

Answer

0 to 1

If three class data set contains the 1/3 records of each of the three class the this dataset has the -

Answer

“Highest impurity”, “1”

Attribute with highest Information Gain is choosen as a root node

Answer

True

Validation Dataset is also called as…..

Answer

“Testing dataset”

_____ method is more suitable when dataset is small

Answer

“K-Fold cross validation”

The Accuracy of model is better in K-fold cross validation method as compare to Hold out method

Answer

“TRUE”

Categorization of all cities in a country into different groups based on their similarity is an example of -

Answer

UnSupervised Learning

Catregorization of Cities in a country into Metro City and Non Metro City is an example of-

Answer

Supervised Learning

Devlopement of Driverless Car is more suited to-

Answer

Reinforcement Learning

One vs One Multiclass Strate1gy does

Answer

n(n-1)/2 Comparisons

Complexity of One vs All Multiclass strategy is

Answer

O(n)

Multiple Regression has -

Answer

Multiple input and multiple out variables

Output of Regression is -

Answer

Numeric

In n-cross validation for Dataset with p records, the maximum value of n can be-

Answer

p

The number of iterations in apriori ___

Answer

increases with the size of the maximum frequent set

Frequent item sets is

Answer

Superset of both closed frequent item sets and maximal frequent item sets