Question #1 of 7

Question ID: 1208622

Big data is *most* likely to suffer from low:

A) velocity.

B) variety.

C) veracity.

Explanation

Big data is defined as data with *high* volume, velocity, and variety. Big data often suffers from low veracity, because it can contain a high percentage of meaningless data.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #2 of 7

Question ID: 1208619

An executive describes her company's "low latency, multiple terabyte" requirements for managing Big Data. To which characteristics of Big Data is the executive referring?

A) Velocity and variety.

B) Volume and variety.

C) Volume and velocity.

Explanation

Big Data may be characterized by its volume (the amount of data available), velocity (the speed at which data are communicated), and variety (degrees of structure in which data exist). "Terabyte" is a measure of volume. "Latency" refers to velocity.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #3 of 7

Question ID: 1208625

When evaluating the fit of a machine learning algorithm, it is *most* accurate to state that:

A) accuracy is the ratio of correctly predicted positive classes to all predicted positive classes.

B) recall is the ratio of correctly predicted positive classes to all actual positive classes.

C) precision is the percentage of correctly predicted classes out of total predictions.

Explanation

Recall (also called sensitivity) is the ratio of correctly predicted positive classes to all actual positive classes. *Precision* is the ratio of correctly predicted positive classes to all predicted positive classes. *Accuracy* is the percentage of correctly predicted classes out of total predictions.

(Study Session 3, Module 8.3, LOS 8.g)

Related Material

SchweserNotes - Book 1

Question #4 of 7

Question ID: 1208621

A government decides it will privatize vehicle registrations if the province's auto insurance companies can record and maintain ownership titles using distributed ledger technology. This application of distributed ledger technology is *best* characterized as:

A) blockchain.

B) tokenization.

C) smart contracts.

Explanation

Tokenization refers to maintaining ownership records for physical assets on a distributed ledger. This might, but would not necessarily, use a blockchain, which is a subcategory of distributed ledgers. Smart contracts are computerized agreements designed to automatically carry out certain actions if defined conditions are met.

(Study Session 3, Module 8.1, LOS 8.e)

Related Material

SchweserNotes - Book 1

Question #5 of 7

Question ID: 1208624

Under which of these conditions is a machine learning model said to be underfit?

A) The model treats true parameters as noise.

B) The model identifies spurious relationships.

C) The input data are not labelled.

Explanation

Underfitting describes a machine learning model that is not complex enough to describe the data it is meant to analyze. An underfit model treats true parameters as noise and fails to identify the actual patterns and relationships. A model that is overfit (too complex) will tend to identify spurious relationships in the data. Labelling of input data is related to the use of supervised or unsupervised machine learning techniques.

(Study Session 3, Module 8.3, LOS 8.d)

Related Material

SchweserNotes - Book 1

Question #6 of 7

Question ID: 1208620

Which of the following uses of data is *most accurately* described as curation?

A) An analyst adjusts daily stock index data from two countries for their different market holidays.

B) A data technician accesses an offsite archive to retrieve data that has been stored there.

C) An investor creates a word cloud from financial analysts' recent research reports about a company.

Explanation

Curation is ensuring the quality of data, for example by adjusting for bad or missing data. Word clouds are a visualization technique. Moving data from a storage medium to where they are needed is referred to as transfer.

(Study Session 3, Module 8.1, LOS 8.a)

Related Material

SchweserNotes - Book 1

Question #7 of 7

Question ID: 1208623

In big data projects, data exploration is *least* likely to encompass:

A) feature selection.

B) feature engineering.

C) feature design.

Explanation

Data exploration encompasses exploratory data analysis, feature selection, and feature engineering.

(Study Session 3, Module 8.2, LOS 8.c)

Related Material

SchweserNotes - Book 1