1. The four V's of big data are Volume, Velocity, Variety and Veracity.

Which of these four V's is applicable when we talk about data continuously changing and expanding?

* Veracity
* Variety
* **Velocity**
* Volume

Explanation:

Volume refers to the incredible amount of data we currently generate.

Velocity refers to data continuously being added and things change very rapidly.

Variety refers to the fact that there often is not one type of data but many types (text, image, audit trails) exist and are often combined.

Veracity refers to the fact that you cannot be completely sure that the data is fully accurate.

----------------------------------------------------------------------------------------------------------------

1. When we talk about replay, we mean the process where…

* we start from a process model and generate behavior, e.g. traces.
* we start from event data and generate a process model, e.g. a Petri net.
* **we start from both a process model and a collection of observed behavior, e.g. traces, and compare these.**

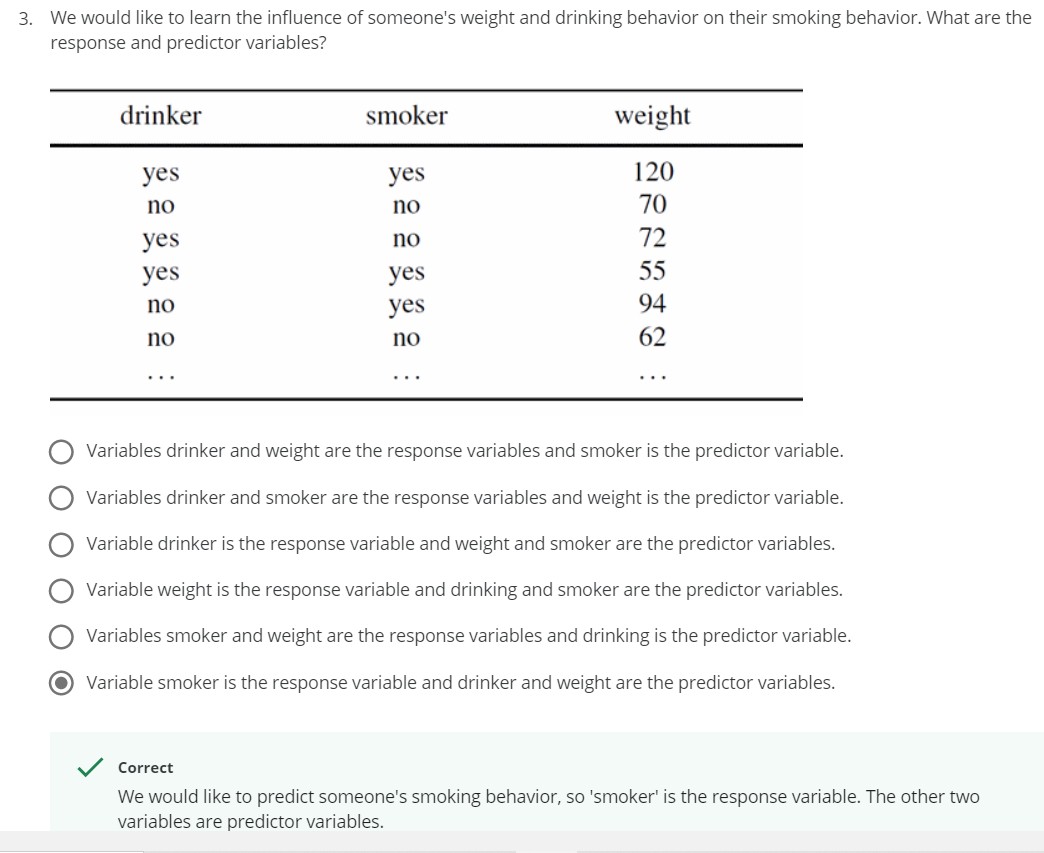
Explanation:

Play-Out is when we start from a process model and generate behavior, e.g. traces.

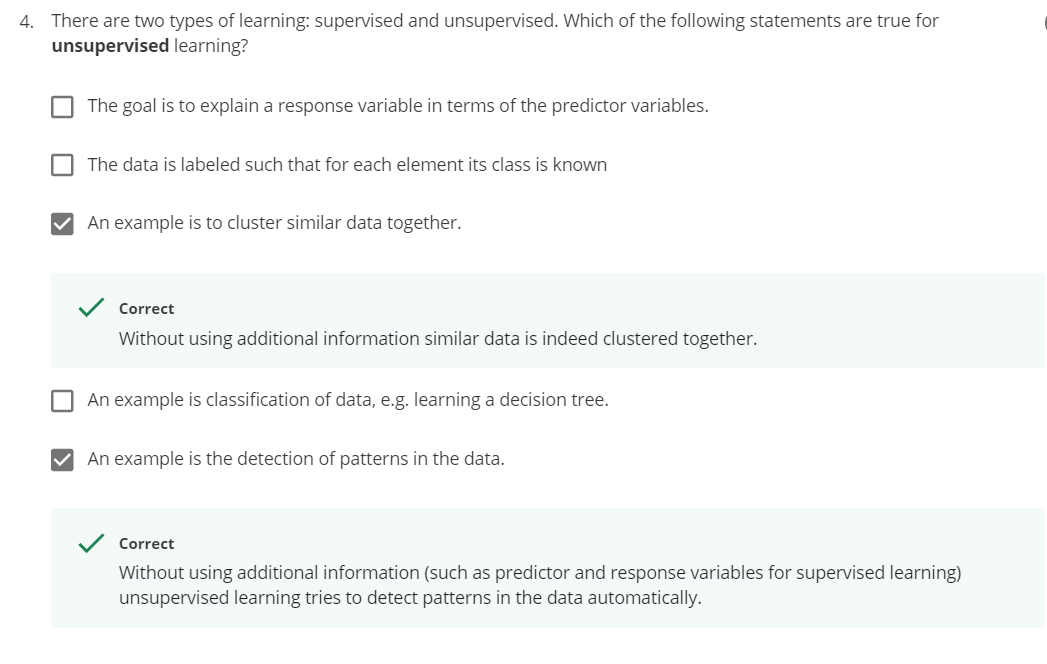
Play-In is when we start from event data and generate a process model, e.g. a Petri net.

Replay is when we start from both a process model and a collection of observed behavior, e.g. traces, and compare these by replaying the traces on the process model.

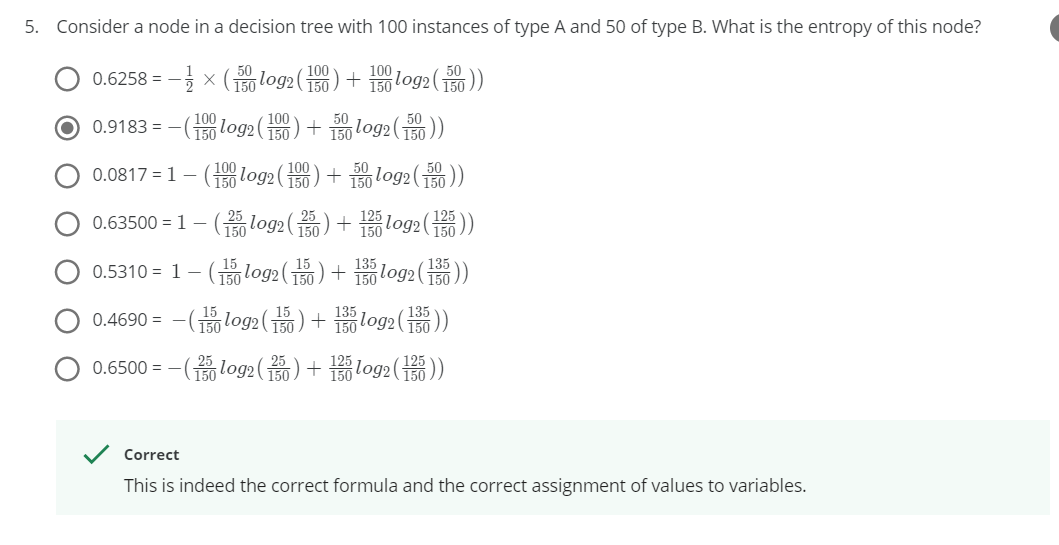
----------------------------------------------------------------------------------------------------------------



------------------------------------------------------------------------------------------------------------

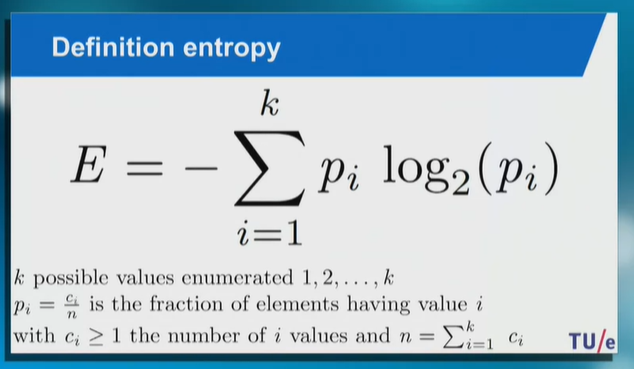


----------------------------------------------------------------------------------------------------------------



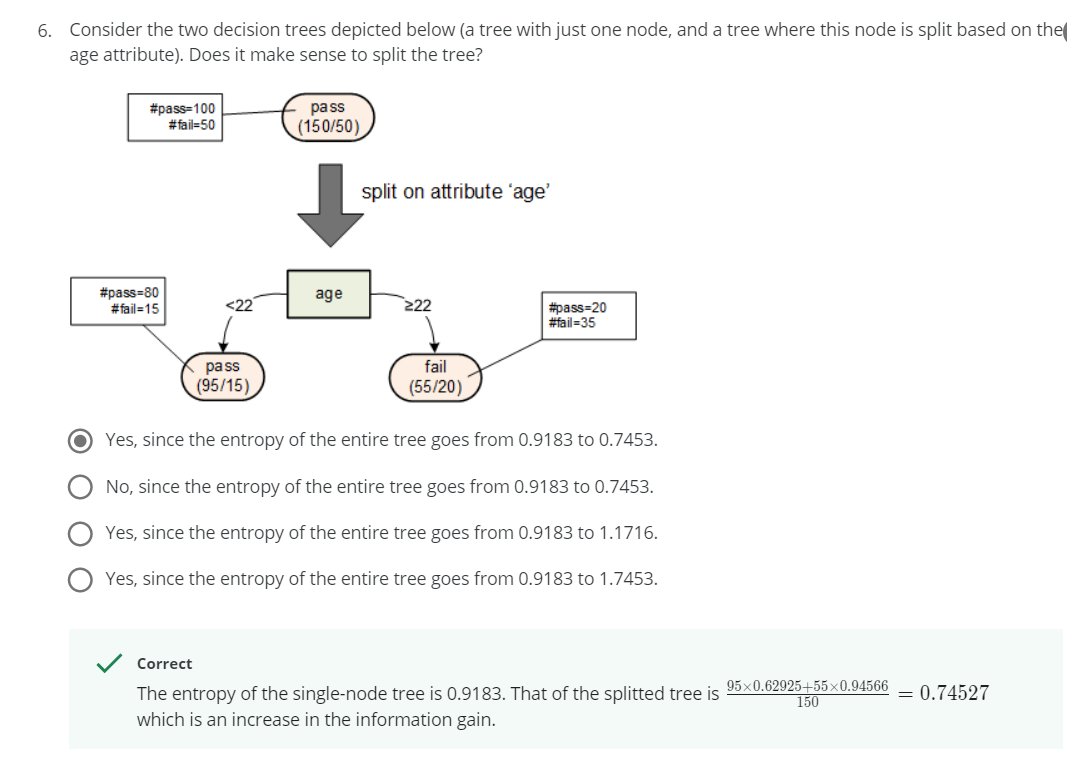
Explanation:

**Entropy** is a measure of the randomness in the information being processed.



E = - (100/150 log(100/150) + 50/150 log(50/150)) = -(-0.389-0.528) = 0.9183

----------------------------------------------------------------------------------------------------------------



Explanation:

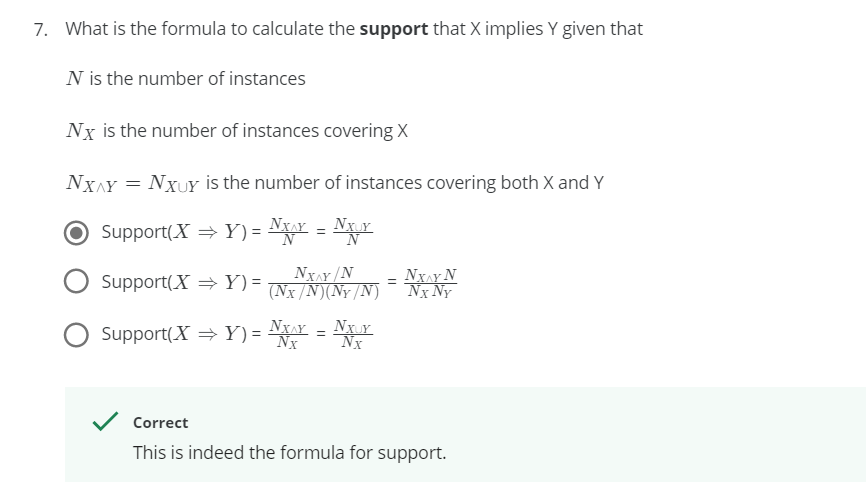
For one node => E = 0.9183

After split: E1 = - (80/95 log(80/95) + 15/95 log(15/95)) = - (-0.208-0.420) = 0.62925

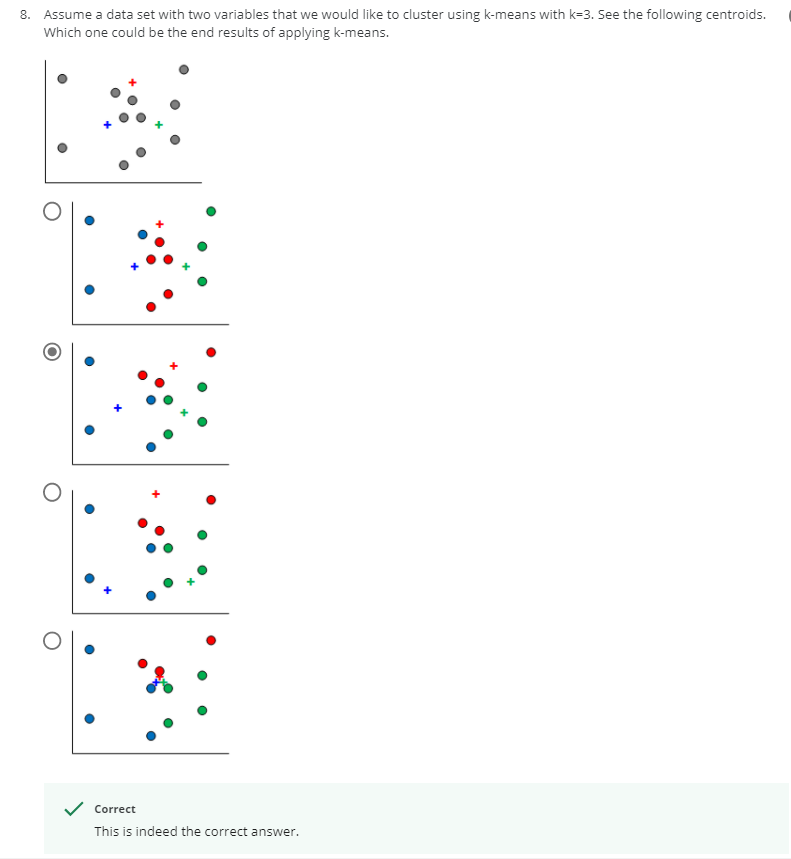
E2 = - (20/55 log(20/55) + 35/55 log(35/55)) = - (-0.5307-0.41495) = 0.94566

Splitted tree entropy = (95\*E1 + 55\*E2)/150 = 0.74527

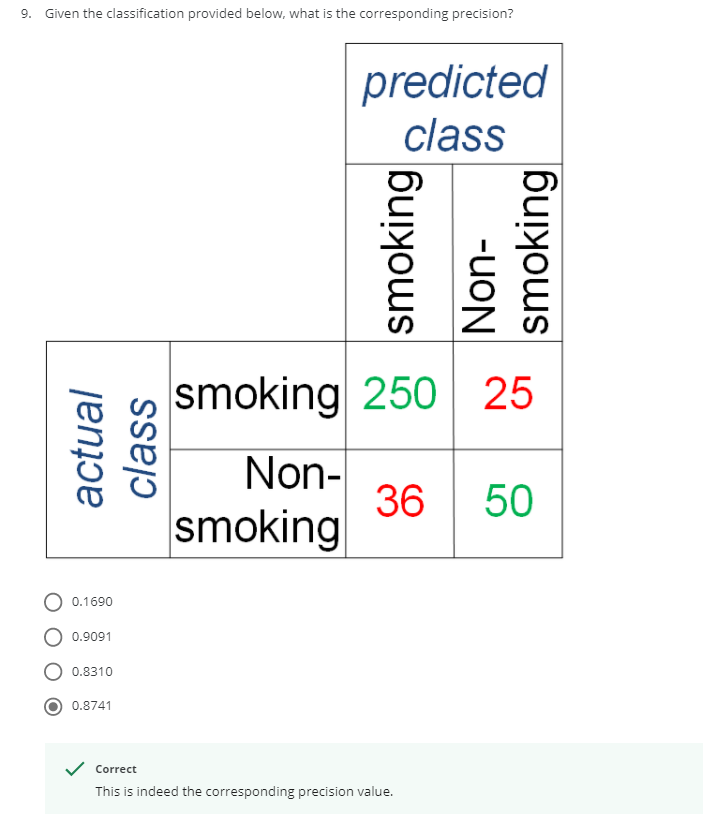
----------------------------------------------------------------------------------------------------------------



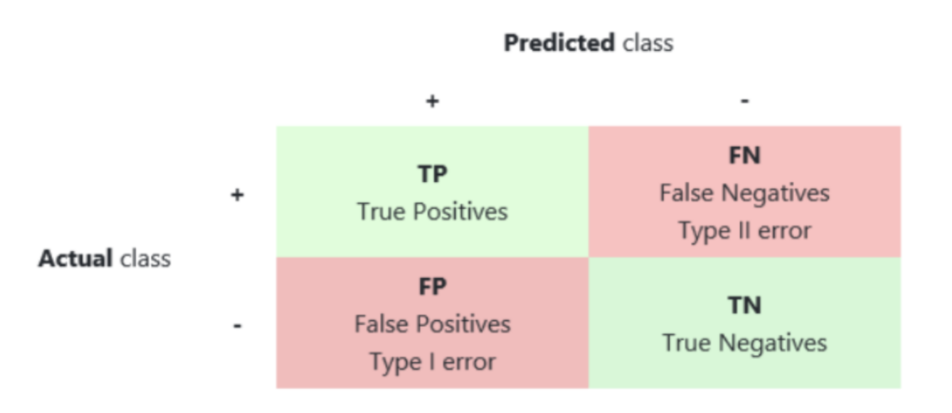
----------------------------------------------------------------------------------------------------------------



----------------------------------------------------------------------------------------------------------------



Explanation:



Precision = TP / (TP+FP) = 250 / (250+36) = 0.8741

Recall = TP / (TP+FN) = 250 / (250+25) = 0.9091

Error = (FP + FN) / (TP + TN + FP + FN) = (36+25) / (250+50+36+25) = 0.1690

Accuracy = (TP + TN) / (TP + TN + FP + FN) = (250+50) / (250+50+36+25) = 0.8310

----------------------------------------------------------------------------------------------------------------

