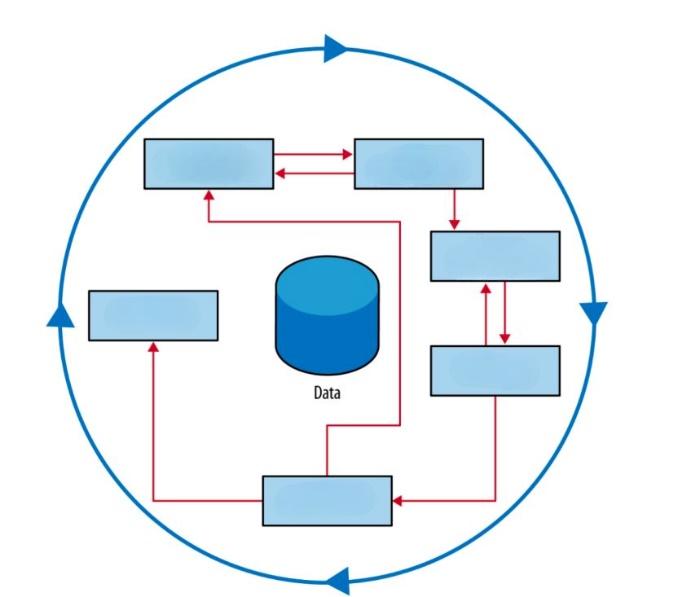
Trial Mid Term

1. Place the items in the correct order.



Problem/Business Understanding

Data Understanding

Data Preparation

Modeling

Evaluation

Deployment

1. In a housing dataset, what type of attribute is the number of bedrooms in a house?

Question 2Answer

a.continuous

**b.discrete**

c.nominal

d.none

e.ordinal

1. What type of [data](https://moodle.uacs.edu.mk/mod/folder/view.php?id=50854) is found in spreadsheets like Google Sheets and Excel?

Question 3Answer

a.unstructured [data](https://moodle.uacs.edu.mk/mod/folder/view.php?id=50854)

**b.structured**[**data**](https://moodle.uacs.edu.mk/mod/folder/view.php?id=50854)

c.ordinal [data](https://moodle.uacs.edu.mk/mod/folder/view.php?id=50854)

d.semi-structured [data](https://moodle.uacs.edu.mk/mod/folder/view.php?id=50854)

1. What are some common examples of unstructured data?

Question 4Answer

a.word documents

b.images

c.videos

d.relational databases

e.spreadsheets

f.XML files

1. A credit card fraud detection model is initially trained on transaction data from 2021, where fraudulent activities mostly involved small, unusual purchases from new merchants. However, as fraudsters adapt their tactics, the types of fraudulent transactions shift to larger purchases or repeated transactions from known merchants in 2023, causing the model to become less effective at identifying fraud and leading to an increase in undetected fraudulent transactions.

The example is an illustration of:

Question 5Answer

a.clustering

b.sampling bias

**c.data drift**

d.none

e.data normalization

1. What can be used to detect outliers in a dataset?

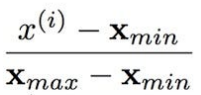
**a.boxplot**

**b.none**

**c.z-score**

**d.scaling**

1. The following formula is used to transform data in the range [0,1]



1. **True**
2. False
3. Choose all the named feature selection methods:

Question 8Answer

**a.cutting methods**

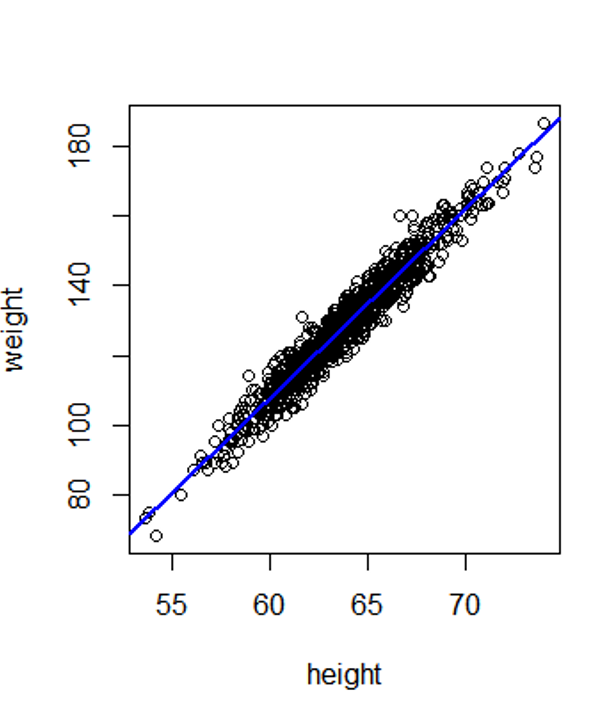
**b.filter methods**

**c.embedded methods**

**d.wrapper methods**

**e.normalization methods**

1. The variables weight and height represented on the Figure have:



a.negative correlation

**b.positive correlation**

c.no correlation

1. What is the mode?

[1,1,1,2,3,4,5,5,6]

Question 10Answer

a.3

b.6

**c.1**

d.None

e.3.11

1. What is the name of the Artificial Intelligence that can perform any task a human can do?

Question 11Answer

**a.artificial general intelligence (AGI)**

b.artificial narrow intelligence (ANI)

c.None

d.artificial custom intelligence (ACI)

1. An agent is learning if it improves its performance after making observations about the world. When the agent is a computer, we call it Answer machine learning
2. What type of learning occurs when the agent receives rewards for gathering coins and faces penalties for losing a life in a game like Super Mario?



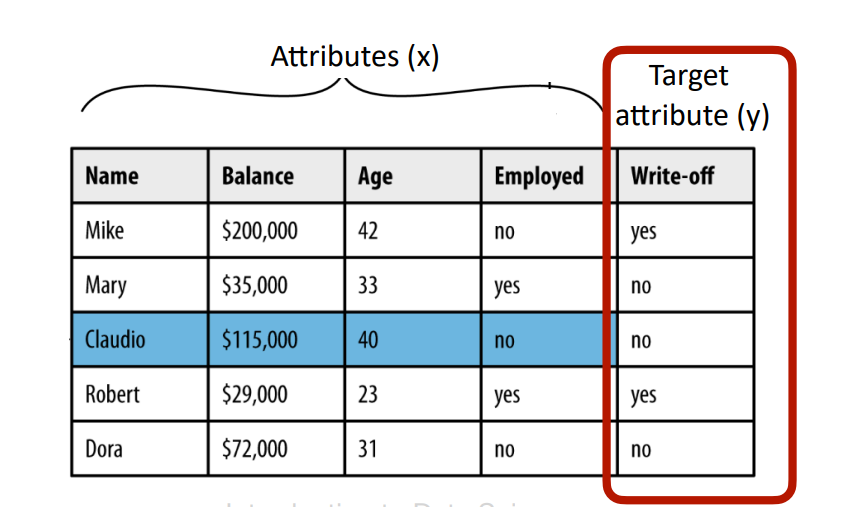
a.Supervised Learning

b.Semi-supervised Learning

**c.Reinforcement Learning**

d.Unsupervised Learning

1. The following task is:



a.multiclass classification

b.regression

**c.binary classification**

d.clustering

1. Models that are overfitting tend to have:

Question 15Answer

a.None

**b.high variance**

c.high bias

d.low variance and low bias

1. We say that the model generalizes well if it accurately predicts the outputs of the training set.

Question 16Answer

True

**False**

1. What is the name of the subset used to tune the hyperparameters of a model?

Question 17Answer

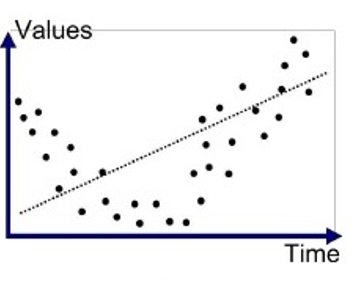
a.training set

b.hyperset

**c.validation set**

d.test set

1. With respect to the data, the linear model in the following picture depicts:

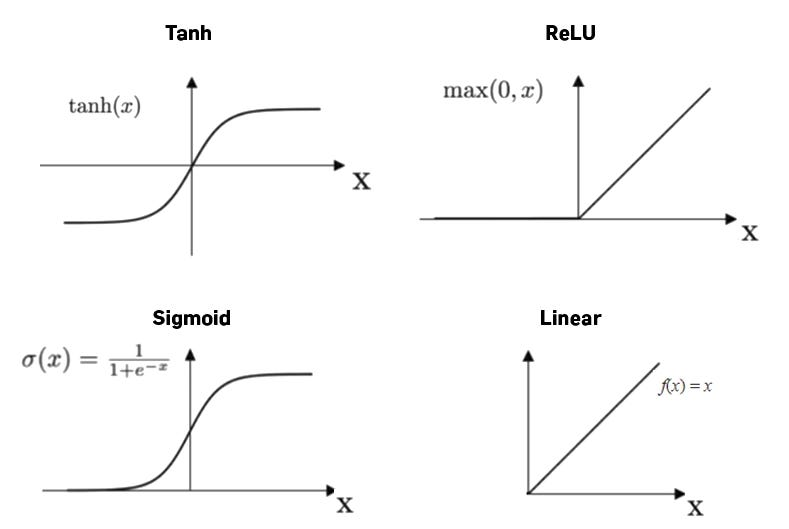


a.neither underfitting nor overfitting

b.overfitting

**c.underfitting**

1. Which function is used by the logistic regression:



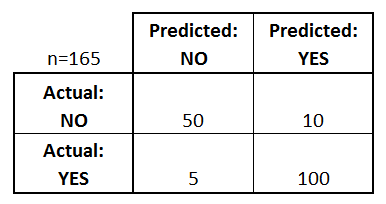
a.Tanh

b.ReLU

**c.Sigmoid**

d.Linear

1. Using the following confusion matrix, calculate the accuracy of the model:



a.105/110

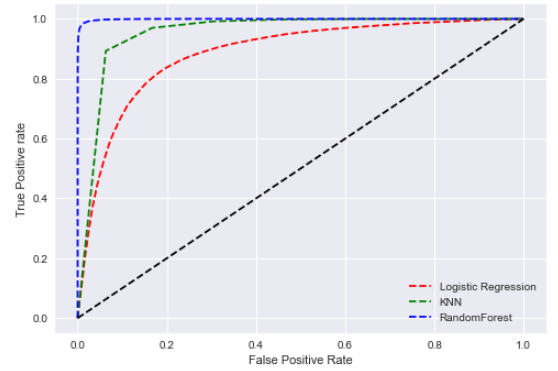
**b.150/165**

c.105/165

d.None

e.110/165

1. Which of the models in the Figure, has the best performance in terms of AUC (Area Under the Curve)?



a.Random Classifier (Black)

b.Logistic Regression (Red)

c.kNN (Green)

d.None

**e.Random Forest (Blue)**

1. Which of the following methods can be used to estimate the coefficients of a logistic regression model?

Question 22Answer

**a.Gradient Descent**

**b.None**

**c.Maximum Likelihood Estimation**

**d.Least Squares**

**e.Unsupervised Learning**