DATA SCIENCE (CDA)
CLASS ASSESSMENT 2 (UNITS 3 AND 4) MODEL A

1. Which kind of task is this?
"Determine what party a person votes for from the information revealed by their tweets"
a) Classification.
b) Regression.
c) Association
d) Clustering

2. Which kind of task is this?
"Determine what combinations of pathologies do old people have "
a) Classification.
b) Correlation.
c) Association
d) Clustering

3. Which kind of task is this?
"Determine what kinds of books I read, according to their genre, length, language and many other features"
a) Classification.
b) Regression.
c) Correlation
d) Clustering

4. Which kind of task is this?
"Determine the money that was spent in each building built in 1998 from its characteristics (location, square metres, construction style, materials, etc.)"
a) Classification.
b) Regression.
c) Correlation
d) Clustering

5. What is the k in k-nearest neighbours (kNN)?
a) The number of groups, as in k-means.
b) The number of nearest neighbours to compare with.
c) The number of layers, as in ANN.
d) The number of kernels, as in SVM.

6. Which of the following can NOT be used for regression?
a) Linear regression.
b) Logistic regression.
c) Neural networks.
d) Non-linear regression.

7. Given the following exact relation between variables: $x_1 = 3.2x_2 - 5.2$ and $x_3 = 5.4x_4 + 2.2$.
a) The correlation between x_1 and x_2 is higher than the correlation between x_3 and x_4.
b) The correlation between x_1 and x_2 is equal to the correlation between x_3 and x_4.
c) The correlation between x_1 and x_2 is lower than the correlation between x_3 and x_4.
d) We cannot know the correlation of these variables, only the slope between them.

8. When should we use cross-validation?
a) Always, it comes by default with many libraries.
b) When we have a small number of examples.
c) When we have a large number of examples.
d) Never, it breaks the golden rule of evaluation.

9. When can we safely discard a binary classifier?
a) Only when its location in the ROC space is both below and right of any other classifier.
b) Only when its location in the ROC space is either below or right of any other classifier.
c) Only when its location in the ROC space is not on the convex hull (pareto front) formed by all classifiers.
d) Only when its location in the ROC space is both below and left of any other classifier.

10. What is collaborative filtering?
a) The recommendations are produced by observing the user's most similar items.
b) The recommendations are produced by observing the preferences of similar users.
c) The recommendations are produced by observing the characteristics of the items (shape, price, colour, etc.).
d) The recommendations are produced by observing the characteristics of the users (age, gender, etc.).

ASSESSMENT
Answer Sheet (MODEL A)

Surname:	**Name:**
Group in English: ☒	

In the following table, circle the correct answer for each question.

Question	**Answer**			
1	**a**	b	c	d
2	a	b	**c**	d
3	a	b	c	**d**
4	a	**b**	c	d
5	a	**b**	c	d
6	a	**b**	c	d
7	a	**b**	c	d
8	a	**b**	c	d
9	a	b	**c**	d
10	a	**b**	c	d

The result will be calculated by the statistical correction formula:

$$(\text{Right} - \text{Wrong}/3) \times 1$$

which discounts the probability of getting a right answer by chance on a question with four possibilities.

The mark is between 0 and 10.

Remember that this assessment is just 10% of the final qualification for the course.