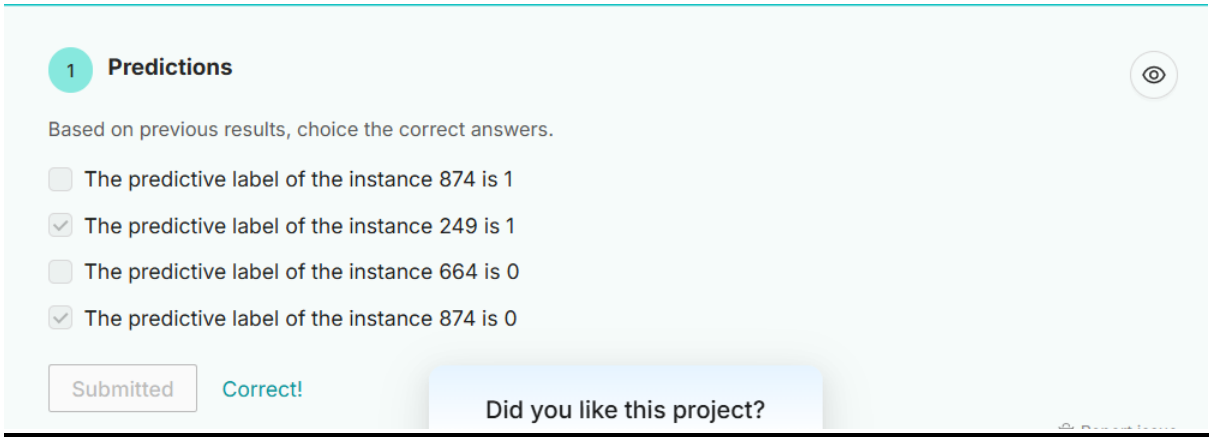
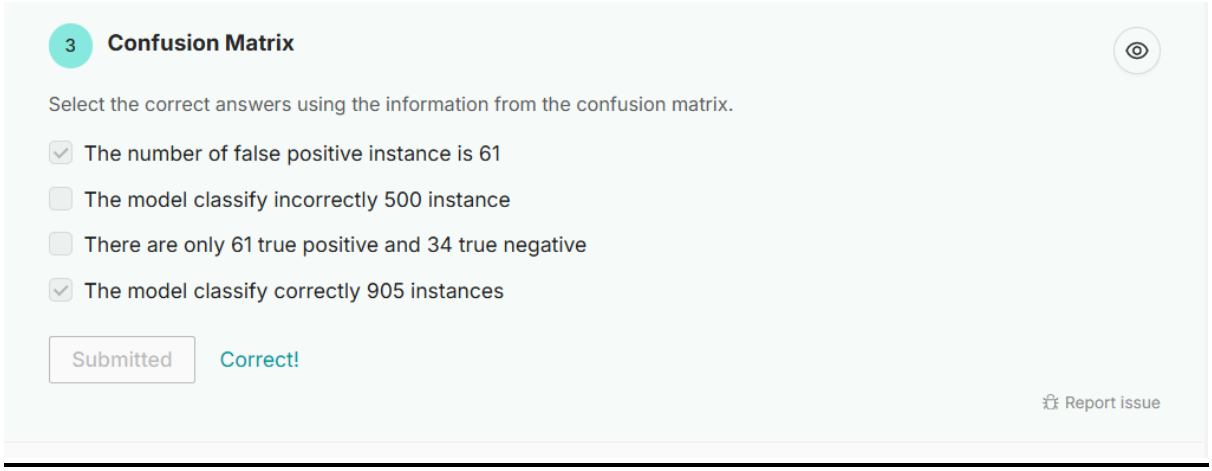
DATAWARS PROJECTS

Learning area: Classification

Shrinivasan M

Project 1: Introduction to Classification





4 Create the response variable based on the columns Trump and Clinton

Separate the target and the features into two variables and create the response variable based on the columns Trump and Clinton.

Select the correct code to complete this task

> *There could be more than just one correct answer.*

- [ ] `X=df.drop(['trump', 'clinton'],axis=1)`
- [ ] `y=df.trump`
- [x] `y=np.where(df.trump>df.clinton,1,0)`
- [x] `X=df[['minority','bachelor']]`

Submitted Correct!

Report issue

5 Decision Tree classifier

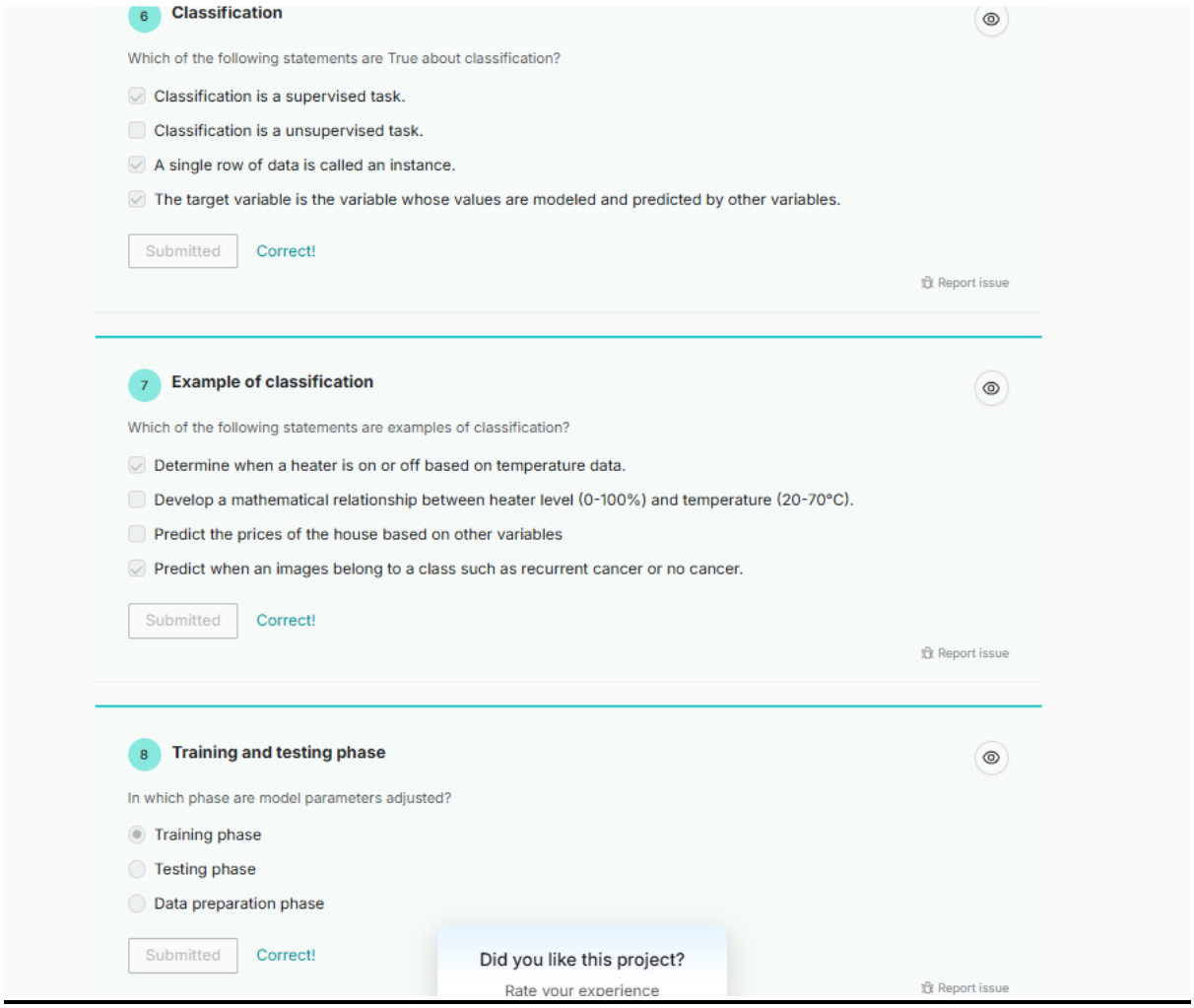
Initialize a Decision Tree classifier (name the model clf) and fit on the data with a random_state: 42 and max_depth : 3 (the maximum depth of our decision tree using the max_depth parameter).

Calculate the accuracy score of the training dataset. Select the correct answer.

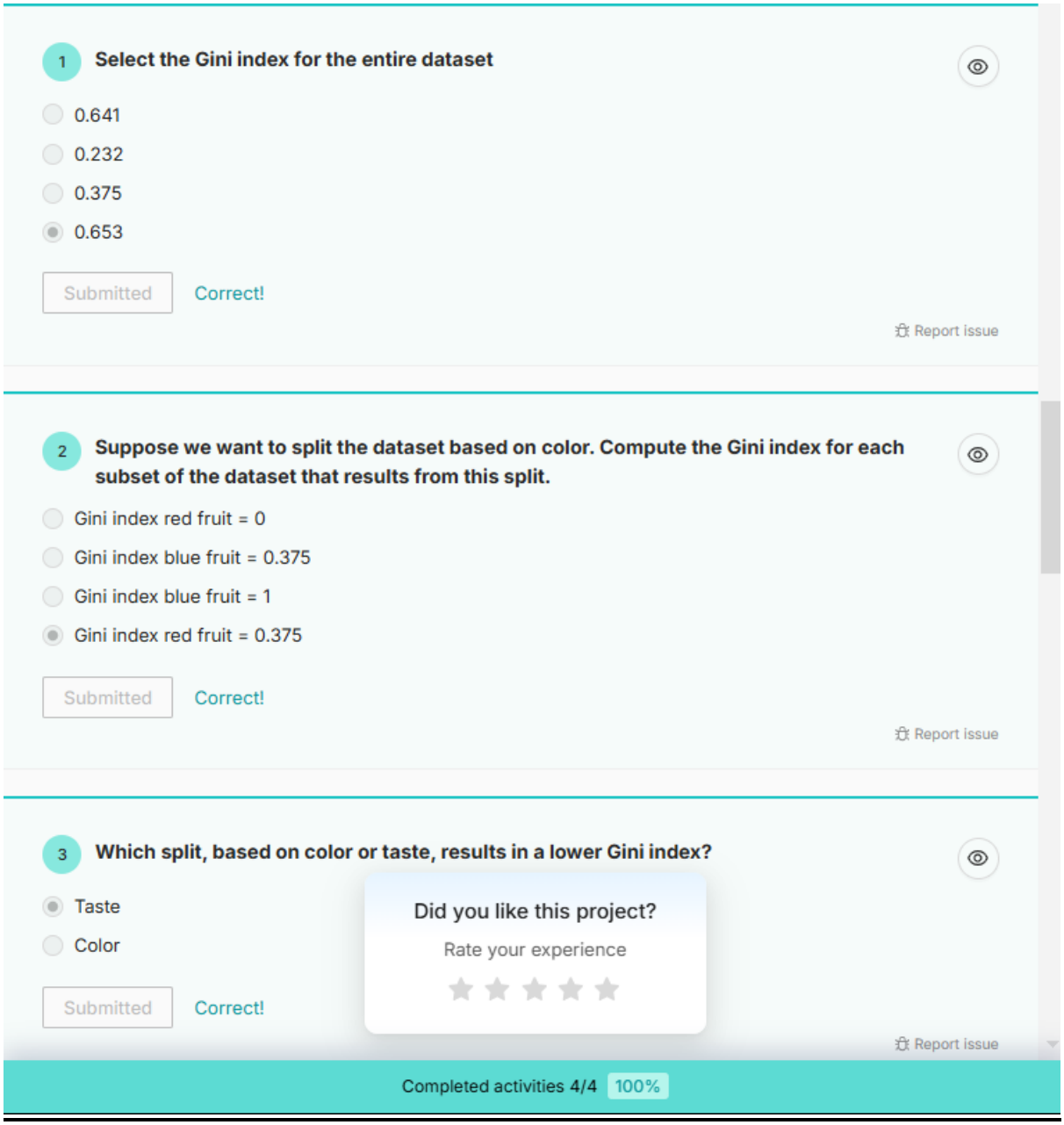
- 0.901
- (•) 0.908
- 0.801
- 0.808

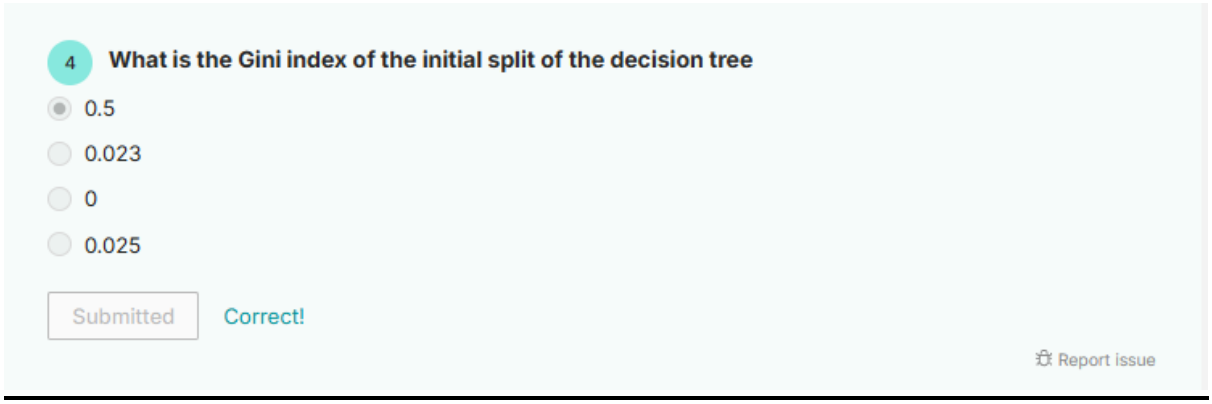
Submitted Correct!

Report issue



Project 2: Coding decision tree from scratch





Project 3: K-Nearest Neighbor algorithm for classification

1 **Performance evaluation**

What is accuracy of training and testing data?

Round to two decimal places

- [ ] The accuracy of the test set is 0.80
- [x] The accuracy of the test set is 0.90
- [ ] The accuracy of the train set is 0.78
- [x] The accuracy of the train set is 0.93

Submitted Correct!

Report issue

2 Test time versus train time

k-NN algorithm does more computation on test time rather than train time.

- No
- Yes

Submitted Correct!

Report issue

3 KNN uses

Which of the following option is true about KNN algorithm?

- It can be used for classification
- It can be used for regression
- It can be used in both classification and regression

Submitted Correct!

Report issue

4 Assumptions

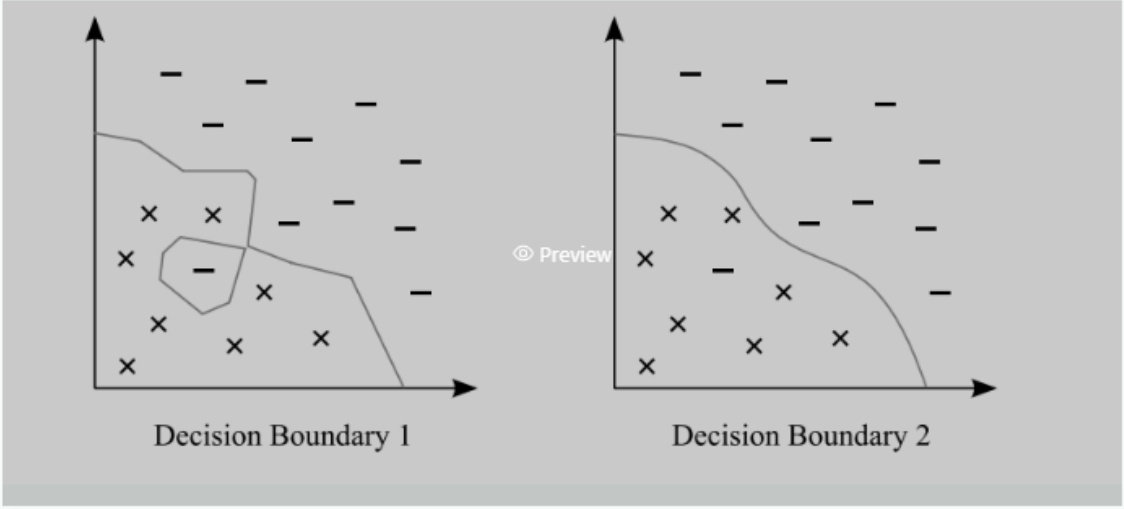
Which of the following statement is true about KNN algorithm ?

- [x] KNN works well with a small number of input variables.
- [ ] k-NN makes assumptions about the functional form of the problem being solved.
- [x] KNN performs much better if all of the data have the same scale.

Submitted Correct!

5 **Decision Boundaries**

The Figure below illustrates decision boundaries for two nearest-neighbour classifiers. Determine which one of the boundaries belongs to the 1-nearest neighbour classifier.



- Decision boundary 2
- Decision boundary 1

Submitted Correct!