

# TERM 3 <br> GRADE 11 - PROJECT 

## Python

## TEACHER'S GUIDELINES



Picture 1


Picture 2

الإلا مـارات الـعـربـــة المتـحــــة


## IMPORTANT NOTICE

Dear teachers, in the case that your school is facing any software issues with Python or PyCharm, please ensure that you create a ticket by completing the online form which was sent by operations.

The link for the online form is as follows:
https://forms.office.com/Pages/ResponsePage.aspx?id=ZN eq0arBUuqA TbXZ4iiiz-35SJZDP1KqLNm-zZZCCdUQjBUMzkwQ0RLRFNNVVhHSENKWDFDMIFIOC4u

Please make the students do the project as and when the concepts are covered while teaching. The following table indicates which project questions to look at for each week, in order complete the entire project on time (this has also been colour coded on the Teacher Guidance work plan):

| Teaching Week | Project Question to work on |
| :---: | :---: |
| WEEK 5 | $\mathbf{1 , 2 , 7 A ^ { * }}$ |
| WEEK 6 | 3, 4, 5, 6, 7B-7G |
|  | * 7A is to be worked on in advance - so all images |
| are prepared for the beginning of week 6 |  |

There can be two ways to maintain the students' assessment work. As per the National Agenda objectives for the UAE vision 2021 for a sustainable environment, and keeping the 'Ministry without papers' initiative, softcopy is strongly recommended.

- Softcopy (electronic)
- Here you will maintain an electronic portfolio (e-portfolio) for each student. An electronic portfolio is a collection of students work completely as softcopy.
- The teacher will collect all student work as softcopy. This should be maintained to produce an evidence for task completion.
- Hardcopy (printed)
- Here you will print the project task and provide each student with a copy. Make the students fill the columns as per the project requirement.
- For any practical (coding) question, a student will type the code using the required software. They will save a copy of their code and the output generated as a screenshot. This screenshot can be printed or can be maintained as a softcopy file.
- The teacher will collect the hardcopy and the screenshot file (softcopy or hardcopy). This should be maintained to produce an evidence for task completion.

الإلا مـارات اللعـربـــة المتـحــــة


## PROJECT OBJECTIVE

You will be applying your knowledge from the Artificial Intelligence Project, covering the following student learning outcomes (SLO's):

- apply programming concepts learned to solve real-world problems.
- explain the basic principles of artificial intelligence and machine learning.
- develop code to classify images of handwritten digits using artificial intelligence and machine learning.
- analyse the performance of machine-learning algorithms.


## EQUIPMENT REQUIREMENTS

Laptop or Computer, Camera, Paper, Black Marker, Python 3.X, PyCharm / Notepad / Notepad++.

## PROJECT INTRODUCTION

To complete the project, you need to perform each of the following:

1. You will answer several questions on 'artificial intelligence' and how it could help the UAE. You will then visually demonstrate how to teach a computer to find out if an image is an orange or a banana using training and testing code.
2. An image is stored as a list of lists. Using the letter ' $A$ ' example, you will show how the letter ' H ', could be represented digitally, and how it can then be represented in a single list in Python.
3. The SKLEARN dataset module in Python contains images of digits. You will answer questions about this dataset, and how we can use Python list functions with the dataset.
4. The SKLEARN dataset module contains a number functions that you will import. You will briefly describe what each function does.
5. To test the machine we use known classified samples. You will answer questions from the code, and describe classification and confusion matrix reports.
6. To test our own images of digits, we need to use some functions of the SCIKIT-IMAGE module beforehand. You will answer questions from the SCIKIT-IMAGE code.
7. You will test the code using 20 of your own handwritten images of digits. You will generate the code to classify all 20 images. You will develop the code to generate your own classification report and confusion matrix for your results. You will answer some questions from your confusion matrix results.

Finally, you will explain and give an advantage or disadvantage of Artificial Intelligence.

## QUESTION GUIDELINES

1. Please fill in the areas highlighted.
2. Type and execute your code using PyCharm software (or any other Python editor).
3. Show your code and output to your teacher.

In the case where you are not able to show your code and output to your teacher:

- save your program.
- take a print screen of your code and an output of your code.
- save your print screen file.
- submit the saved file as a hardcopy or softcopy to your teacher.


## TEACHER RECOMMENDATIONS

- Answers may vary. We request the teachers to take professional judgement for marking the project. It is common knowledge that there is no one way of coding, and that coding styles differ. But it is recommended to use the format/structure given for expected output generation for assessment.
- Do not copy and paste the code provided in the teacher's guide, as the PyCharm editor may show errors due to formatting or otherwise. The code provided should not be given to the students.
- Maintain every student work for evidence.

الإبـارات العـريـيـة المتـحــدة
وزارة التــربـيــة والـتـعــــيـم

## PROJECT - WORK PLAN



An image is stored as a list of lists.
A. Using the shown letter ' $A$ ' example, show how the letters ' $\mathbf{U}$ ' and ' $\mathbf{E}$ ', could be represented, for the UAE.

Please shade the boxes in black.
B. Show how your letters could be digitally stored in a single final list called imageU and iageE
A. Show how the letters 'U' and 'E' can be represented (Hint, for each try on a blank piece of paper first, on an $8 \times 8$ grid )
Answers will vary. Please use your professional judgement.

## Letter U:



## Letter A:



Letter E:

listel $=[1,0,0,0,0,0,0,1]$ liste2 $=[1,0,1,1,1,1,1,1]$ liste3 $=[1,0,1,1,1,1,1,1]$ IistE4 $=[1,0,0,0,0,0,0,1]$ IistE5 $=[1,0,1,1,1,1,1,1]$
IistE6 $=[1,0,1,1,1,1,1,1]$ listE6 $=[1,0,1,1,1,1,1,1]$
listE7 $=[1,0,0,0,0,0,0,1]$ Iiste8 $=[1,1,1,1,1,1,1,1]$
imagee = [listel, liste2, liste3, liste4, liste5, liste6, liste7, liste8]

| \# | Work Steps | Step Completion \& Values | Rem arks |
| :---: | :---: | :---: | :---: |
| 3 | The SKLEARN dataset module contains images of digits. <br> A. For a machine to learn, which of the following number of images in its dataset would provide a more correct answer? Why? <br> a) 10 images <br> b) 100 images <br> c) 1000 images <br> B. Fill in the function name that returns the length of the list 'images'. What is the list length when running the code? <br> C. Write the code to put the first 1000 elements of the list 'images' into a sublist. <br> D. To prepare the data for machine learning we will use the first 1000 elements in the SKLEARN dataset module for training. What will we use the remaining elements in the SKLEARN dataset module for? | A. Answer: <br> 1000 images would provide a more correct answer. <br> Reason: If there is a larger number of images for the machine to learn, then the system will have more samples to make predictions from, and the answer will be more accurate. <br> B. Fill in the function name that returns the length of the list 'images'. ```from sklearn.datasets import load_digits digits = load_digits() images = digits.images print("Total number of images:", len (images))``` <br> What is the list length returned when running this code? <br> Answer: Total number ©f images: 1797 <br> C. Answer: $\text { imagesSublist }=\quad \text { images[0:1000] }$ <br> D. Answer: <br> For testing. <br> There are 1797 images in total that we can use, so the first 1000 to train the system, the remainder to see how accurate the system is for the remaining 797. This will give us an idea of how accurate the system is, and the accuracies can be seen on the confusion matrix report. | - <br> - - <br> $\because$ |

الإلمـارات العـربـيـة المتـحــــة
وزارة التــربـيـة والـتـعـلـيـم

4 The SKLEARN dataset module contains a number functions that you will import.
A. Briefly describe what the following functions do:

- load_digits,
- SVC,
- classification_report
- confusion_matrix
B. 'training_samples', 'training_targets', 'testing_samples' and 'testing_targets' are all types of?


## Code Extract:

```
from sklearn.datasets import load digits
from sklearn.svm import SVC
from sklearn.metrics import classification_report,
            confusion_matrix
```

digits = load_digits()
images = digits.images
target $=$ digits.target
data $=$ images.reshape ( 1797,64 ) $)$
classifier = SVC(gamma=0.001)
training_samples = data[0:1000]
training_targets = target[0:1000]
testing_samples = data[1000:1797]
testing_targets = target[1000:1797]
A. The load_digits function:

It will load the dataset of 1,797 digits with their correct labels
The SVC function: Short for Support Vector. It will carry out the classification work.

The classification_report and confusion_matrix functions:
These functions will check how good your code will be at figuring out the digit from the image of that digit
B. Answer:
list (or sublist is also acceptable)

الإلمـارات الـعـربـيـة المتـحــــة
وزارة التتـربـيــة واتـتـعـلـيـم

5 To test the machine we use known classified samples. Answer the following questions from the code:
A. What does
classifier.fit() do when we pass run this function with the training_samples and training_targets lists?
B. What does classifier.predict() do when we pass run this function with the testing_samples list?
C. Look at the following classification report, displayed from running the code.

Which did the computer classify easier, digit '3' or digit '6'?
D. Look at the following confusion matrix report displayed. The fifth row highlighted in the matrix gives the confusion result for digit 4.

What can the information in this row tell us?

## Code Extract:

classifier.fit(training_samples, training_targets)
predicted = classifier.predict(testing_samples)
print("Classification report for classifier")
print(classification_report(testing_targets,
predicted))
print("Confusion matrix")
print(confusion_matrix(testing_targets, predicted))
A. Answer:

The classifier.fit() function links the training samples to the training targets.
B. Answer:

The classifier.predict() passes an item in the testing_samples list and tries to predict what the digit number is. It is like checking how well a student knows the subject, where classifier.predict() is the 'taking the exam' part.
C. Classification report:

Answer: Digit 6.

D. Confusion matrix report:


Answer:
This The fifth row of this matrix tells us that there were 79 testing questions which had the answer 4 that the computer also classified as 4. It got them right. However, there was four cases where a 4 was confused with a 9 .

| 6 | To test our own images of digits, <br> we need to use some functions <br> of the SCIKIT-IMAGE module <br> beforehand. |
| :--- | :--- |
| Answer the following questions |  |
| from the code: |  |

A. What does imread () do when we pass run this function with the 'testImage.png' image file, and the word 'True'?
B. The functions resize () and reshape () are performed to resize and reshape the images. Why do we need to do this?
C. Fill in the blanks in this sentence.

## Code Extract:

```
from skimage.transform import resize
from skimage import io
```

```
img = io.imread(testImage.png, True)
img = resize(img, (8,8), mode='reflect')
img = (1-img) * 16
vector = img.reshape((1,64))
print(classifier.predict(vector))
```

A. Answer:
imread() function loads an image from a file into the variable img and changes it from colour to black and white
B. Answer:

The images you take maybe different in size from the images taken by your classmates or found on the Internet. The images must be sized to $8 \times 8$ because this is what you used in training.
C. 'vector' is a long _list $\qquad$ of pixels. The function
classifier.predict() will then predict the $\qquad$ class of the image.

7 Test the code using your own images of digits.
A. Handwrite the digits 0 to 4 four times, and save as seperate images.
B. To read all 20 images, we need to ensure that the code looks at them.
A. On a separate page, using a black marker write digits 0 to 4 like below, four times. This will make a total of 20 images of digits. (* Hint, use different handwriting, or handwriting from a different student)


Take a picture of the above filled in table of digits and use the SnippingTool in Windows,to 'crop' your images into 20 individual images of each digit.

Name each of the 20 images appropriately, for example for the fourth image of your digit 3 could be "image3-4.png".


## Handwritten <br> on paper



Snipping Tool used to crop each image, for example:
image2-2.png

Save all 20 image separate files into a folder.
B. Drag all 20 image files into your Python project. Show your teacher an image of at least one of your own handwritten digits in your Python project.

The code extract below classifies two of the digits (0 \& 1):

```
#images to classify
allimages0=["image0-1.png", "image0-2.png",
    "image0-3.png", "image0-4.png"]
allimages1=[ "image1-1.png", "image1-2.png",
    "image1-3.png", "image1-4.png"]
for thisimage in allimages0 + allimages1:
```

Develop the code to test the machine on all the images you saved. Write the code needed to classify all 20 of your own images of digits:

```
allimages0=["image0-1.png","image0-2.png",
    "image0-3.png", "image0-4.png"]
allimages1=["image1-1.png","image1-2.png",
    "image1-3.png", "image1-4.png"]
allimages2=["image2-1.png","image2-2.png",
    "image2-3.png", "image2-4.png"]
allimages3=["image3-1.png","image3-2.png",
    "image3-3.png", "image3-4.png"]
allimages4=["image4-1.png","image4-2.png",
    "image4-3.png", "image4-4.png"]
for thisimage in allimages0 + allimages1 + allimages2 + allimages3
+ allimages4:
    img = io.imread(thisimage, True)
    img = resize(img, (8,8), mode='reflect')
    img = (1-img) * 16
    vector = img.reshape((1,64))
    print("\n =a= == == == == == == == == == == == == =")
    print("Image", thisimage)
    print(classifier.predict(vector))
```

C. In the code, generate the classification report and confusion matrix for all your images.

## PLEASE SPEAK TO YOUR TEACHER FOR GUIDANCE.

C. Write here the code that generates the classification report and confusion matrix for your own images of digits:

```
answers=[0,0,0,0,1,1,1,1,2,2,2,2,3,3,3,3,4,4,4,4]
predicted = []
for thisimage in allimages0 + allimages1 + allimages2 + allimages3
+ allimages4:
    img = io.imread(thisimage, True)
    img = resize(img, (8,8), mode='reflect')
    img = (1-img) * 16
    vector = img.reshape((1,64))
    print("\n =a= == == == == == == == == == == == == =")
    print("Image", thisimage)
    print(classifier.predict(vector))
    predicted.append(classifier.predict(vector))
    print("\n == == == == == == == == == == == == == =")
#Creating a classification report and confusion matrix for the
results of your images
import numpy
predicted = numpy.array(predicted)
answers = numpy.array(answers)
print("Classification report for classifier")
print(classification_report(answers, predicted))
print("Confusion matrix")
print(confusion_matrix(answers, predicted))
```

D. Run the code for all your digit images. Fill in your own confusion matrix for the 20 images of digits.
E. Which of your digits did the computer find easiest and hardest to classify correctly?
F. Take a screenshot of your completed code, and the generated output.
D. Fill in your own confusion matrix.

E. Easiest digit: $\square$ Hardest Digit: $\qquad$ \#
F. Generated output Screenshot:

G. Answer:

This is an open-ended question, where answers will vary. Please take professional judgement for this answer.

## RUBRIC \& MARKING RUBRIC GUIDANCE

1) Print the rubric for each student and fill in the student details.
2) Check the work for each criteria and circle the cell which you feel is appropriate with respect to the criteria completion.

When you look at the student work, please take a professional judgement on the points which is required to be awarded.
3) Enter the marks for each of the criterias in the space given underneath, and add together to calculate the total out of 50 .
www.almanahj.com
$\qquad$

| Criteria | Excellent | Very Good | Good | Inadequate | Un satisfactory | 0 |
| :---: | :---: | :---: | :---: | :---: | :---: | :---: |
| 1. Applied programming concepts learned to solve realworld problems. | -Clear application of list data structure and control structures in the code <br> - Clear coding standards have been applied with the correct logic and syntax and compiles with no errors. 910 points | - Good application of list data structure and control structures in the code -Good coding standards have been applied with the correct logic and syntax and compiles with no errors. 7-8 points | -Satisfactory application of list data structure and control structures in the code <br> -Satisfactory coding standards have been applied with the satisfactory logic and syntax and compiles with minor warnings 5-6 points | -Weak application of list data structure and control structures in the code <br> -Weak coding standards have been applied with the incorrect logic and syntax and compiles with errors. 3-4 points | - No application of list data structure and control structures in the code <br> - No coding standards have been applied with the incorrect logic and syntax and numerous errors. 1-2 points | No effort made |
| 2. Explained the basic principles of machine learning. | -Clear machine learning definition and letters are represented in 1 \& 0's in lists correctly. <br> -Machine learning diagrams are filled in completely and correctly.9-10 points | -Good machine learning definition and letters are represented in $1 \& 0$ 's in lists correctly. <br> -Machine learning diagrams are mostly filled in and correct.7-8 points | - Satisfactory machine learning definition and letters are mostly represented in 1 \& 0's in lists satisfactorily. <br> - Machine learning diagrams are mostly filled in and mostly correct.5-6 points | -Weak machine learning definition and letters are represented in 1 \& 0's in lists weakly. <br> -Machine learning diagrams are partially filled in, but not correctly.3-4 points | - No correct definition of machine learning and letters are not represented in 1 \& 0's in lists correctly. <br> - Machine learning diagrams are not filled nor correctly.1-2 points | No effort made |
| 3. Develop code to classify images of handwritten digits using AI. | - Clear AI definition and UAE Al reason given. <br> -Discussion of AI is clear. <br> - Clear understanding and development of code using the SKLEARN, SCIKITIMAGE and classifier code functions 13-15 points | -Good AI definition and UAE Al reason given. <br> -Discussion of Al is good. <br> - Good understanding and development of code using the SKLEARN, SCIKITIMAGE and classifier code functions 10-12 points | - Satisfactory Al definition and UAE AI reason given. <br> -Discussion of Al is satisfactory. <br> - Satisfactory understanding and development of code using the SKLEARN, SCIKITIMAGE and classifier code functions7-9 points | -Weak Al definition and UAE Al reason given. <br> -Discussion of Al is weak. <br> -Weak understanding and development of code using the SKLEARN, SCIKITIMAGE and classifier code functions 4-6 points | - No correct AI definition and UAE AI reason given. <br> -Discussion of Al is unsatisfactory. <br> - No understanding and development of code functions 1-3 points | No effort made |
| 4. Analyse and evaluate the performance of machinelearning algorithms. | -Clear understanding of classification and confusion matrix reports. <br> -Classification report and confusion matrix are run successfully. <br> - Able to clearly analyse the performance of the machine learning based on the report generated.13-15 points | - Good understanding of classification and confusion matrix reports. <br> - Classification report and confusion matrix are run generally with success. <br> - Able to analyse well the performance of the machine learning based on the report generated.10-12 points | -Satisfactory understanding of classification and confusion matrix reports. <br> -Classification report and confusion matrix are run satisfactorily. <br> -Can analyse satisfactorily the performance of the machine learning based on the report generated.7-9 points | -Weak understanding of classification and confusion matrix reports. <br> -Classification report and confusion matrix are not run satisfactorily. <br> - Unable to analyse satisfactorily the performance of the machine learning based on the report generated.4-6 points | - No understanding of classification and confusion matrix reports. <br> -Classification report and confusion matrix are not run. <br> - No analysis made regarding the performance of the machine learning based on the report generated.1-3 points | No effort made |
|  |  | + | + | + | + |  |
|  |  |  |  | TOTAL: |  | / 50 |

