Python Final Quiz - Updated

Total points 45/50

INT: Intellipaat DS-AI
Intellipaat DS-AI

The respondent's email (**grv08singh@gmail.com**) was recorded on submission of this form.

Full Name: *

Gaurav Singh

✓ 1.What will be the result of the following code? * 1/1

```
import numpy as np

data = np.array([5, 15, 25, 35, 45])
scaled = (data - np.min(data)) / (np.max(data) - np.min(data))
print(scaled)
```

- ○ Data scaled between -1 and 1
- ○ Data centered around 0 with unit variance
- ◉ Data scaled between 0 and 1 ✓
- ○ Data without any change
- ○ Other:

✓ 2.Suppose two events X and Y are such that $P(X)$ = 0.5 P(X)=0.5, $P(Y)$ = 0.3 P(Y)=0.3, and $P(X \cap Y)$ = 0.1 P(X∩Y)=0.1. What is $P(X \cup Y)$ P(X∪Y)? *1/1

- (●) 0.7 ✓
- ( ) 0.8
- ( ) 0.9
- ( ) 0.6
- ( ) Other: _____

✓ 3.Which of the following is a key assumption of one-way ANOVA? * 1/1

- ( ) The predictor variables are uncorrelated
- ( ) The response variable is categorical
- (●) The variances within each group are equal ✓
- ( ) The groups must have the same number of observations
- ( ) Other: _____

✗ 4.In the dataset provided, the target variable is Any Previous Loan. *0/1

Based on the dataset, compute the **entropy of the target variable Any Previous Loan**.

Job	Education	Any Previous Loan	Salary	Housing Loan
Business	Schooling	No	Medium	No
Business	PG degree	No	Medium	Yes
Salaried	UG degree	No	High	No
Business	PG degree	Yes	Low	Yes
Business	Schooling	No	Low	Yes
Salaried	UG degree	Yes	Medium	Yes
Salaried	Schooling	No	Low	Yes
Business	UG degree	Yes	Medium	No
Salaried	UG degree	No	Low	No
Business	PG degree	Yes	Low	Yes
Business	PG degree	No	Low	Yes
Salaried	UG degree	No	High	Yes
Business	Schooling	Yes	High	Yes
Business	PG degree	No	Medium	No
Salaried	PG degree	No	High	No
Business	Schooling	Yes	High	Yes
Salaried	Schooling	Yes	Medium	No

- ○ 0.971
- ○ 0.934
- ○ 1.000

0.845

Other: 0.977 ✗

Correct answer

0.971

✓ 5. A company has two departments: * 1/1

Department A:

Employees = 60

Average Salary = ₹45,000

Standard Deviation = ₹4,000

Department B:

Employees = 40

Average Salary = ₹50,000

Standard Deviation = ₹5,000

What is the combined standard deviation (rounded to 2 decimal places)?

- ₹4,46.53 ✓
- ₹4,24.26
- ₹4,60.15
- ₹4,56.78

✓ 6.What will be the correlation coefficient between two variables related as: *1/1

$y=-2x$

- ○ 1
- ◉ -1 ✓
- ○ 0
- ○ 2

✓ 7. What is the probability that an employee is Female, given that the employee has quit? *1/1

Gender	Quit: Yes	Quit: No
Male	0.38	0.12
Female	0.40	0.10

- ○ 0.45
- ○ 0.48
- ◉ 0.51 ✓
- ○ 0.61

✓ 8. Which sampling method divides the population into groups and then randomly selects entire groups? *1/1

- ○ Stratified sampling
- ◉ Cluster sampling ✓
- ○ Systematic sampling
- ○ Convenience sampling

✓ 9. You are analyzing survey results to check if there's an association between **education level** (UG, PG) and **job preference** (IT, Finance, Teaching).
Which test should you use? *1/1

- ○ ANOVA
- ○ Paired t-test
- ◉ Chi-square test ✓
- ○ Z-test

✓ 10.You're studying the effect of three different diets on weight loss. Which statistical test would be appropriate to determine if there's a significant difference in **average weight loss** across the groups? *1/1

- ○ t-test
- ◉ ANOVA ✓
- ○ Chi-square
- ○ Kruskal-Wallis test

✓ 11. What can be concluded from the output if the printed p-value is 0.0002? *1/1

```
import pandas as pd
from scipy import stats

batch_A = [82, 85, 90, 88, 86]
batch_B = [78, 76, 80, 79, 77]
batch_C = [91, 89, 92, 94, 90]

f_stat, p_value = stats.f_oneway(batch_A, batch_B, batch_C)
print("P-value:", p_value)
```

- ○ There is no significant difference in means across batches
- ○ Only batch_A and batch_C are significantly different
- ◉ At least one batch mean is significantly different from the others ✓
- ○ ANOVA test is not suitable here

✓ 12. Which of the following encoding techniques is most suitable for converting an **ordinal categorical variable** (e.g., "Low", "Medium", "High")? *1/1

- ○ One-hot encoding
- ◉ Label encoding ✓
- ○ Binary encoding
- ○ Hashing encoding

✓ 13. Given the restaurant dataset, which variable is **most likely to require dummy encoding** before fitting a linear regression model? *1/1

Variable	Description
Price	price of a meal
Food	measure of quality
Decor	measure of decor of the restaurant
Service	measure of quality of service
East	dummy variable indicating if the restaurant is east or west of 5th avenue

- ○ Price
- ○ Food
- ○ Decor
- ◉ East ✓

✓ 14. Given the confusion matrix: * 1/1

	Predicted: Defective	Predicted: Not Defective
Actual: Defective	30	20
Actual: Non-Defective	120	830

What is the F1 Score of the model?

- 0.20
- 0.15
- 0.17
- 0.30 ✓

✓ 15. Let the vectors be: * 1/1

$$\vec{v_1} = \begin{bmatrix} 3 \\ 4 \end{bmatrix}, \quad \vec{v_2} = \begin{bmatrix} 6 \\ 8 \end{bmatrix}$$

What is the cosine similarity between the two vectors?

- 1.0 ✓
- 0.8
- 0.6
- 0.5

✓ 16.Which of the following statements is TRUE regarding the Binomial, Poisson, and Normal distributions? *1/1

- ○ Binomial and Poisson are continuous, Normal is discrete
- ◉ Binomial and Poisson become approximately Normal under certain conditions ✓
- ○ Normal distribution is always skewed
- ○ Poisson distribution has two parameters: mean and standard deviation

✓ 17. What will be the output of Following code : * 1/1

```
import numpy as np

arr2d = np.array([[11, 12, 13, 14],
                  [21, 22, 23, 24],
                  [31, 32, 33, 34],
                  [41, 42, 43, 44]])

print(arr2d[1:3, 1:], arr2d[::-1, 0], arr2d[:2, ::-1])
```

```
[[22 23 24]
 [32 33 34]]
[41 31 21 11]
[[14 13 12 11]
 [24 23 22 21]]
```

◉ A ✓

```
[[22 23 24]
 [32 33 34]]
[41 31 21 11]
[[24 23 22 21]
 [14 13 12 11]]
```

○ B

```
[[12 13 14]
 [22 23 24]]
[11 21 31 41]
[[24 23 22 21]
 [14 13 12 11]]
```

○ C

```
[[22 23 24]
 [32 33 34]]
[11 21 31 41]
[[24 23 22 21]
 [14 13 12 11]]
```

○ D

✓ 18.Consider the following statements about Python functions and closures: *1/1

1. A closure in Python is a function object that has access to variables in its lexical scope, even when the function is called outside that scope.

2. The global keyword is used to define variables that are accessible across all Python files in a project.

3. The nonlocal keyword allows assignment to variables in the nearest enclosing scope (excluding globals).

4. Lambda functions in Python can contain multiple expressions.

- (●) Only 1 and 3 ✓
- ( ) Only 1 and 2
- ( ) Only 1, 3, and 4
- ( ) All of the above

✓ 19. Which of the following best describes the key idea behind stacking in ensemble learning? *1/1

- ( ) Combining weak learners by averaging their predictions
- ( ) Using majority voting among base classifiers to improve accuracy
- (●) Training a meta-model to combine predictions of multiple base models ✓
- ( ) Applying boosting to reduce model bias

✓ 20.You have perform PCA on a dataset with 5 features. The explained variance ratio for the first 3 principal components is: *1/1

```
PC1 = 0.55, PC2 = 0.30, PC3 = 0.10
```

How many principal components should you retain to explain at least 95% of the variance in the data?

- ○ 5
- ○ 4
- ◉ 3 ✓
- ○ 2

✓ 21. You have a categorical feature City with 1000 unique city names. You apply one-hot encoding without any dimensionality reduction. What are potential **drawbacks** of this approach in a machine learning? *1/1

- ○ It will lead to multicollinearity among encoded columns
- ◉ It will increase sparsity and memory usage, potentially causing overfitting ✓
- ○ One-hot encoding will fail if the feature is not ordinal
- ○ One-hot encoding will reduce model complexity

✓ 22. You're analyzing a time series and notice significant autocorrelation in both the **ACF (Autocorrelation Function)** and **PACF (Partial Autocorrelation Function)** plots for multiple lags. Based on this, which model is most appropriate to consider? *1/1

- ○ MA model
- ○ AR model
- ◉ ARMA model ✓
- ○ Naive model

✓ 23. Which of the following statements best distinguishes **XGBoost** from traditional **Gradient Boosting**? *1/1

- ○ XGBoost uses bagging instead of boosting
- ○ XGBoost builds trees in parallel, unlike gradient boosting
- ◉ XGBoost includes regularization to reduce overfitting ✓
- ○ XGBoost uses entropy as the default loss function

✓ 24. While analyzing multicollinearity in a dataset using a **correlation heatmap**, you find two variables with a correlation coefficient of **0.95**. What is the most appropriate next step? *1/1

- ◉ Use one of the variables and drop the other ✓
- ○ Use both variables since high correlation improves accuracy
- ○ Normalize both variables and keep them
- ○ Convert both variables to categorical

✓ 25. In a class of 100 students: *1/1

60 students passed Math,

70 students passed Science,

50 students passed both Math and Science.

What is the probability that a student passed Science, given that they passed Math?

- ○ 0.50
- ○ 0.70
- ◉ 0.83 ✓
- ○ 0.60

✓ 26. What will be the output of the following code snippet? * 1/1

```
sentence = "Data Science is powerful"
words = sentence.split()
result = ''

for word in words:
    if len(word) % 2 == 0:
        result += word[::-2] + ' '
    else:
        result += word[::2] + ' '

print(result)
```

- (●) "Dt cec s pwrfl " ✓
- ( ) "aD ienei swoeful "
- ( ) "Dt cie swoeful "
- ( ) "aD ienei s pwrfl "

✓ 27. Which of the following data types can be used as a key in a Python dictionary? *1/1

- ( ) list
- (●) tuple ✓
- ( ) dictionary
- ( ) set

✗ 28. What will be the **output** of the following code, and **how many vowels** appear in it? *0/1

```
word = "understanding"
output = ''

for i in range(1, len(word), 2):
    if word[i] not in ['a', 'e', 'i', 'o', 'u']:
        output += word[i]

print(output)
```

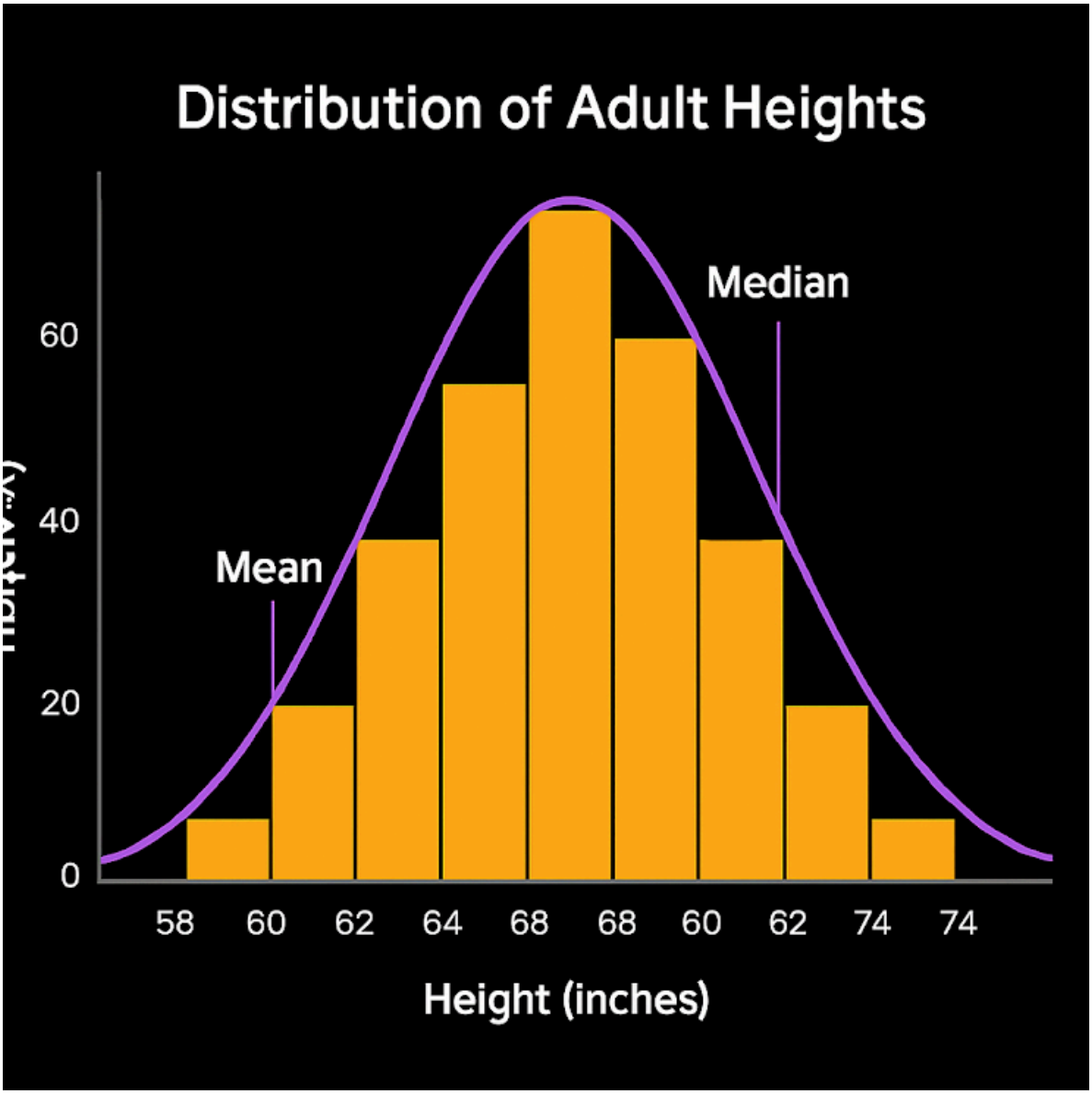
- ○ ntrsnn, Vowels = 0
- ○ ntrsn, Vowels = 1
- ○ ntrsnn, Vowels = 1
- ◉ ndrsnn, Vowels = 0 ✗

Correct answer

- ◉ ntrsnn, Vowels = 0

✗ 29. The following histogram shows the distribution of adult heights in inches. *0/1

Based on the graph, what can you say about the relationship between the mean and the median?



- ○ Mean = Median
- ○ Mean > Median
- ○ Mean < Median
- ◉ Cannot be determined ✗

Correct answer

- ◉ Mean = Median

✓ 30. In a company, 30% of employees work in the tech department. Out of the tech employees, 70% use Python. Among the non-tech employees, only 10% use Python. *1/1

If a randomly selected employee uses Python, what is the probability that they are from the tech department?

- 0.21
- 0.3
- 0.7
- 0.75 ✓

✓ 31. Based on the given dataset of x and y values: * 1/1

x	y
1.15	-4.64
-2.16	2.98
-1.73	1.99
-1.63	1.76
8.79	-22.21
2.76	-8.34
3.46	-9.95

Which of the following best describes the relationship between x and y?

- ○ Monotonic increase
- ○ Monotonic decrease
- ◉ Non-monotonic ✓
- ○ Constant relationship

✓ 32. You are working with the student attendance DataFrame shown on the **left**. You want to remove duplicate rows **based only on the first_name and last_name columns**, keeping the **first occurrence**. *1/1

Which of the following code snippets will result in the table shown on the **right**?

(Refer to the image for table before and after duplication removal)

	first_name	last_name	age	attendance
1	Hari	Ram	24	25
2	Hari	Ram	34	25
3	Ash	Duo	36	57
4	Wasim	Ali	25	62

	first_name	last_name	age	attendance
1	Hari	Ram	24	25
2	Hari	Ram	34	25
3	Ash	Duo	36	57
4	Wasim	Ali	25	62

- (●) df1.drop_duplicates(subset=['first_name', 'last_name'], keep='first') ✓
- ( ) df1.drop_duplicates(subset=['first_name', 'last_name'], keep='last')
- ( ) df1.drop_duplicates()
- ( ) df1.drop(['first_name', 'last_name'])

✓ 33. The Third Quartile(Q3) value is : * 1/1

index	CLAY_COMPOSITION
count	352
mean	7.95
std	0.86
min	5.42
25%	7.44
50%	8.10
75%	8.54
max	10.36

- ○ 7.44
- ◉ 8.54 ✓
- ○ 8.10
- ○ 10.36

✗ 34. What will be the output of the following code? * 0/1

```
x = [10, 20, 30, 40, 50, 60, 70]
y = slice(1, -1, 2)
print(x[y])
```

- (●) [20, 40, 60] ✗
- ( ) [20, 30, 40, 50, 60]
- ( ) [20, 40]
- ( ) IndexError

Correct answer

- (●) [20, 40]

✓ 35. Consider the following dataset with 2 features: * 1/1

Observation	Feature A	Feature B
1	2	8
2	4	10
3	6	12
4	8	14
5	10	16

After applying Principal Component Analysis (PCA) to this dataset:

Which of the following is most likely true about the eigenvalues of the principal components?

- (●) First principal component will have an eigenvalue much larger than the second. ✓
- ( ) Both principal components will have roughly equal eigenvalues.
- ( ) Second principal component will have a much larger eigenvalue.
- ( ) PCA cannot be applied to this dataset.

✓ 36. Which model is most appropriate to forecast the **monthly electricity demand** in a city using historical consumption data and external factors like temperature and economic activity? *1/1

- ( ) KMeans Clustering
- ( ) Random Forest Regressor
- (●) SARIMAX ✓
- ( ) Naive Bayes

✓ 37. Suppose you have two variables, **X** and **Y**, with a correlation coefficient of **0.75**. If **X** is **multiplied by -3** and **Y** is **divided by 2**, what will be the new correlation coefficient? *1/1

- (●) -0.75 ✓
- ( ) 0.75
- ( ) -1.5
- ( ) 0

✓ 38. You are given the following dataset representing the exam scores of 20 students: *1/1

[78, 82, 85, 88, 89, 90, 91, 91, 92, 93, 94, 95, 96, 96, 97, 98, 99, 99, 100, 30]

Based on the data, which of the following is the **correct interpretation of the distribution**?

- ( ) The distribution is symmetric
- ( ) The distribution is right-skewed
- (●) The distribution is left-skewed ✓
- ( ) The distribution is normally distributed

✓ 39. You are working with the *Mall Customer Segmentation Data*, which contains the following columns: *1/1

CustomerID

Age

Annual Income (k$)

Spending Score (1-100)

You are tasked with segmenting the customers into groups based on **Annual Income** and **Spending Score** using **K-Means clustering**. To determine the **optimal number of clusters**, you plot both the **Elbow Method (WCSS)** and **Silhouette Scores** for K = 2 to 10.

Based on the below WCSS and Silhouette values:

K	WCSS	Silhouette Score
2	444.15	0.65
3	267.90	0.71
4	190.65	0.72
5	150.35	0.69
6	120.30	0.64

What is the most appropriate value of K based on both WCSS and Silhouette Score?

- ○ K = 2
- ○ K = 3
- ◉ K = 4 ✓
- ○ K = 6

✗ 40. You are working with a standardized dataset df_scaled and want to determine the optimal number of clusters for K-Means using the **Silhouette Score**.
Which of the following code snippets will achieve this correctly? (more than one option can be correct). *0/1

```
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_score

for k in range(2, 11):
    model = KMeans(n_clusters=k, init='random', n_init=10)
    labels = model.fit_predict(df_scaled)
    print(f"{k}: {silhouette_score(df_scaled, labels):.2f}")
```

A

```
from sklearn.metrics import silhouette_score

model = KMeans(n_clusters=3)
model.fit(df_scaled)
print(silhouette_score(df_scaled, model.labels_))
```

B

```
from sklearn.cluster import KMeans
from sklearn.metrics import silhouette_samples

for k in range(2, 6):
    kmeans = KMeans(n_clusters=k, random_state=42)
    score = silhouette_samples(df_scaled, kmeans.fit(df_scaled))
    print(f"K={k}: {score.mean():.3f}")
```

C ✗

NONE

Correct answer

A

✓ 41. You applied K-Means clustering on a dataset with scaled numerical features and got the following output: *1/1

```
from sklearn.cluster import KMeans
import pandas as pd

df = pd.read_csv("data.csv")  # contains no missing values and is already scaled
kmeans = KMeans(n_clusters=4, random_state=0)
df['cluster'] = kmeans.fit_predict(df)

print(df.groupby('cluster').mean())
```

- ○ It displays the most frequent data point in each cluster to understand centroids.
- ◉ It reveals how each original feature contributes to the centroid of each cluster. ✓
- ○ It helps in selecting the best k by comparing means across clusters.
- ○ It calculates how many data points fall into each cluster.

✓ 42. You have a high-dimensional dataset (X) with 1000 features and 1000 samples. You want to apply K-Means clustering **after reducing dimensionality** using PCA. You run the following code: *1/1

```
from sklearn.decomposition import PCA
from sklearn.cluster import KMeans

pca = PCA(n_components=0.95)
X_reduced = pca.fit_transform(X)

kmeans = KMeans(n_clusters=3, random_state=42)
clusters = kmeans.fit_predict(X_reduced)
```

What does n_components=0.95 mean in this code?

- ○ It selects the top 95 principal components regardless of explained variance.
- ◉ It keeps components until 95% of the total variance in X is explained. ✓
- ○ It reduces the number of features to 95.
- ○ It uses only the first 95 features of the dataset for clustering.

✓ 43. Which of the following statements is **true** about applying Principal Component Analysis (PCA) **before** K-Means clustering? *1/1

- ○ PCA always improves clustering performance by increasing the distance between clusters.
- ◉ PCA can help reduce noise and improve clustering by projecting data into a lower-dimensional space that preserves most of the variance. ✓
- ○ PCA and K-Means are both supervised learning techniques, so applying one before the other is redundant.
- ○ Using PCA before K-Means is discouraged because it removes class labels and can distort cluster structure.

✓ 44. Which of the following is **NOT** a valid assumption or characteristic of K-Means clustering? *1/1

- ○ Clusters are spherical and equally sized in the feature space.
- ○ K-Means tries to minimize the within-cluster sum of squares (inertia).
- ◉ K-Means performs well when clusters have different densities and non-convex shapes. ✓
- ○ K-Means is sensitive to the initial placement of cluster centroids.

✓ 45. Which of the following statements is **true** regarding the differences between Bagging and Boosting in ensemble learning? *1/1

- ○ Boosting builds base learners independently, while bagging builds them sequentially.
- ○ Bagging focuses on reducing bias, while boosting primarily reduces variance.
- ◉ Boosting assigns more weight to misclassified instances in subsequent iterations, while bagging treats all instances equally. ✓
- ○ Bagging usually leads to overfitting, whereas boosting handles it better by training models in parallel.

✓ 46. Consider the following Python code snippet using sklearn: * 1/1

```
from sklearn.tree import DecisionTreeClassifier
from sklearn.datasets import make_classification
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score

X, y = make_classification(n_samples=1000, n_features=20, random_state=42)

X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.3, random_state=42)

model1 = BaggingClassifier(base_estimator=DecisionTreeClassifier(max_depth=1), n_estimators=50)
model2 = AdaBoostClassifier(base_estimator=DecisionTreeClassifier(max_depth=1), n_estimators=50)

model1.fit(X_train, y_train)
model2.fit(X_train, y_train)

pred1 = model1.predict(X_test)
pred2 = model2.predict(X_test)

print("Bagging Accuracy:", accuracy_score(y_test, pred1))
print("Boosting Accuracy:", accuracy_score(y_test, pred2))
```

- ○ Both classifiers will perform almost the same since they use the same base estimator.
- ○ BaggingClassifier will significantly outperform AdaBoostClassifier due to its parallel processing.
- ◉ AdaBoostClassifier will likely outperform BaggingClassifier because it reduces bias by focusing on misclassified instances. ✓
- ○ Both classifiers will fail to learn anything since the base estimator is too weak.

✓ 47. Which of the following statements are **true** about Object-Oriented Programming in Python? (More than one option can be correct). *1/1

- ○ The __init__ method in a class is automatically called when an object is created and acts as a destructor.
- ◉ Encapsulation in Python is achieved by defining class methods as private using double underscores (e.g., __method()), which completely hides them from outside the class. ✓
- ○ NONE
- ○ Polymorphism enables the same function name to be used for different types, supporting method overloading in Python in the traditional sense.

✓ 48. What will be the output of the following Python code? * 1/1

```
class Sample:
    def __init__(self):
        self.value = 5

    def display(self):
        print(self.value)

obj1 = Sample()
obj2 = obj1
obj2.value = 10
obj1.display()
```

- ○ 5
- ◉ 10 ✓
- ○ Error – obj2 has no attribute value
- ○ Nothing will be printed

✓ 49. What will be the output of the following Python code? * 1/1

```
x = [(2, 3), (1, 4), (3, 2)]
x.sort(key=lambda pair: pair[1] - pair[0])
print(x)
```

- (●) [(3, 2), (2, 3), (1, 4)] ✓
- ( ) [(1, 4), (2, 3), (3, 2)]
- ( ) [(1, 4), (3, 2), (2, 3)]
- ( ) [(3, 2), (1, 4), (2, 3)]

✓ 50.Which of the following statements about Python functions is/are **true**? *1/1

1.A function can return multiple values in Python.

2.Variables defined inside a function are accessible outside the function by default.

3.Default arguments are evaluated each time the function is called.

4.A function in Python can be assigned to a variable.

- (●) Only 1 and 4 ✓
- ( ) Only 2 and 3
- ( ) Only 1, 3, and 4
- ( ) Only 1, 2, and 4