

# Examen de Certificación en Ciencia de Datos y ML

Formato: opción múltiple (teoría + código). Secciones de 15–20 min.

Tiempo total sugerido: 10 módulos x 15–20 min = 150–200 min.

## Módulo 1: Fundamentos de Programación en Python (15–20 min)

1. En Python, ¿cuál de las siguientes estructuras de datos es **mutable**?

- A) Tupla
- B) Lista
- C) Cadena
- D) Entero

2. Completa el código para importar la librería NumPy con su alias estándar:

```
import ____ as np
```

- A) numpy
- B) np
- C) pandas
- D) sklearn

3. ¿Cuál de las siguientes opciones es la manera correcta de definir una función en Python?

- A) `function my_func():`
- B) `def my_func:`
- C) `def my_func():`
- D) `func my_func():`

4. ¿Qué imprime el siguiente código?

```
x = [10, 20, 30]  
print(x[1])
```

- A) 10
- B) 20
- C) 30
- D) Error

5. Ordena los pasos para recorrer una lista `nums` e imprimir sus elementos:

- A) 1-2-3
- B) 2-3-1

- C) 3-1-2
- D) 2-1-3

6. Detecta el error en este código:

```
for i in range(5)
    print(i)
```

- A) `range` no existe
  - B) Falta el `:` al final del `for`
  - C) `i` no está definido previamente
  - D) `print` no acepta enteros
7. ¿Qué librería es más utilizada en Python para manipular datos tabulares (tipo Excel/CSV)?
- A) NumPy
  - B) Pandas
  - C) Matplotlib
  - D) Seaborn
8. Completa el código para crear una lista por comprensión con los cuadrados de los números del 1 al 5:

```
squares = [____ for i in range(1,6)]
```

- A) `i**2`
  - B) `i*i`
  - C) `pow(i,2)`
  - D) Todas las anteriores son correctas
9. ¿Cuál es la forma más eficiente en Python de calcular la suma de una lista `x`?
- A) `sum(x)`
  - B) `x.sum()`
  - C) `np.suma(x)`
  - D) `suma(x)`
10. ¿Cuál es el error en este código?

```
import numpy as np
arr = np.array[1,2,3,4]
```

- A) `np.array` no existe
- B) Falta paréntesis, debería ser `np.array([1,2,3,4])`
- C) `arr` no puede contener enteros
- D) `np` debe importarse como `numpy`

**Respuestas Módulo 1:** 1:B, 2:A, 3:C, 4:B, 5:B, 6:B, 7:B, 8:A, 9:A, 10:B.

## Módulo 2: Obtención y Preparación de Datos (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em] Completa el código para leer un archivo CSV en un DataFrame llamado `df`:

11.

```
import pandas as pd
df = pd.____("archivo.csv")
```

A) `load_csvread_csv`

B) `open_csvcsv_reader`

12) ¿Qué método de Pandas devuelve las primeras 5 filas de un DataFrame?

A) `df.first()`

B) `df.top()`

C) `df.head()`

D) `df.begin()`

13. ¿Cuál de las siguientes líneas elimina valores nulos en el DataFrame `df`?

A) `df.remove_nulls()`

B) `df.dropna()`

C) `df.delete_na()`

D) `df.clean()`

14. Completa el código para rellenar valores faltantes con la media en la columna `"edad"`:

```
df["edad"].fillna(____, inplace=True)
```

A) `df.mean()`

B) `df["edad"].mean()`

C) `np.mean()`

D) `mean(df)`

15. Detecta el error en este código:

```
from sklearn.preprocessing import MinMaxScaler
scaler = MinMaxScaler
X_scaled = scaler.fit_transform(X)
```

A) `MinMaxScaler` no existe en `sklearn`

B) Falta paréntesis: `MinMaxScaler()`

C) `fit_transform` no acepta `X`

D) `x` debe ser un array NumPy

16. Completa el código para aplicar **StandardScaler**:

```
from sklearn.preprocessing import ____
scaler = StandardScaler()
X_scaled = scaler.fit_transform(X)
```

- A) sklearn
- B) StandardScaler
- C) preprocessing
- D) pandas

17. ¿Qué opción es correcta para crear variables dummy de la columna "sexo" en df con Pandas?

- A) `pd.get_dummies(df, "sexo")`
- B) `pd.dummies(df["sexo"])`
- C) `pd.get_dummies(df, columns=["sexo"])`
- D) `df.to_dummy("sexo")`

18. Ordena los pasos correctos para aplicar OneHotEncoder de scikit-learn:

- A) 1-2-3
- B) 2-1-3
- C) 3-2-1
- D) 2-3-1

19. ¿Qué sucede si no escalamos los datos antes de usar KNN?

- A) Nada, siempre funciona igual
- B) Una variable con valores grandes dominará la distancia
- C) KNN dejará de entrenar
- D) Los datos se vuelven categóricos

20. Completa el código para normalizar datos en el rango [0,1] usando MinMaxScaler:

```
scaler = MinMaxScaler(feature_range=(0, ____))
```

- A) 10
- B) 100
- C) 1
- D) None

**Respuestas Módulo 2:** 11:B, 12:C, 13:B, 14:B, 15:B, 16:B, 17:C, 18:B, 19:B, 20:C.

### Módulo 3: Análisis Exploratorio de Datos (EDA) (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em]¿Cuál es el objetivo principal del EDA (Exploratory Data Analysis)?

21. A) Predecir con alta precisión  
B) Detectar patrones, relaciones y outliers en los datos  
C) Implementar modelos supervisados  
D) Normalizar siempre los datos

22. Completa el código para obtener estadísticas descriptivas de un DataFrame `df`:

```
df.____()
```

- A) `stats`  
B) `summary`  
C) `describe`  
D) `info`

23. ¿Qué función muestra la información de columnas, tipos de datos y valores nulos en un DataFrame?

- A) `df.types()`  
B) `df.info()`  
C) `df.describe()`  
D) `df.structure()`

24. ¿Cuál es el gráfico más adecuado para ver la distribución de una variable numérica?

- A) Gráfico de barras  
B) Histograma  
C) Gráfico de torta  
D) Heatmap

25. Completa el código para graficar un histograma de la columna `edad` con Pandas:

```
df["edad"].____()
```

- A) `hist()`  
B) `plot_hist()`  
C) `plot.hist()`  
D) `bar()`

26. Detecta el error en este código de Matplotlib:

```
import matplotlib.pyplot as plt
plt.plot.hist(df["edad"])
plt.show()
```

- A) `df` no puede graficarse en Matplotlib
- B) `plt.plot.hist` no existe
- C) Falta importar seaborn
- D) `plt.show()` no funciona con histogramas

27. ¿Qué gráfico se utiliza para analizar la correlación entre variables numéricas?

- A) Pie chart
- B) Boxplot
- C) Heatmap
- D) Barplot

28. Ordena los pasos para graficar un heatmap de correlaciones con Seaborn:

- A) 1-2-3-4
- B) 2-1-3-4
- C) 3-2-1-4
- D) 2-3-1-4

29. Completa el código para graficar un diagrama de dispersión entre `.edad` y `ingresos` con Pandas:

```
df.plot.scatter(x="edad", y="_____")
```

- A) age
- B) ingreso
- C) ingresos
- D) salary

30. ¿Qué indica un boxplot con bigotes muy largos y puntos alejados?

- A) Datos normalizados
- B) Posible presencia de outliers
- C) Correlación lineal alta
- D) Baja varianza

**Respuestas Módulo 3:** 21:B, 22:C, 23:B, 24:B, 25:C, 26:B, 27:C, 28:B, 29:C, 30:B.

## Módulo 4: Inferencia Estadística (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em]¿Qué mide un intervalo de confianza del 95 %?

31. A) Que el 95 % de los datos caen en ese rango
- B) Que el 95 % de los intervalos calculados contendrán el verdadero parámetro poblacional
- C) Que la media siempre está dentro del intervalo
- D) Que los datos siguen distribución normal

32. Completa el código para calcular la media de una columna `edad` en Pandas:

```
media = df["edad"].____()
```

- A) `mean()`
- B) `average()`
- C) `promedio()`
- D) `media()`

33. En una prueba de hipótesis, el p-value indica:

- A) La probabilidad de que  $H_0$  sea verdadera
- B) Evidencia en contra de la hipótesis nula
- C) La media poblacional
- D) La desviación estándar

34. ¿Qué distribución se usa en muestras grandes con varianza conocida?

- A) t de Student
- B) Normal (Z)
- C) Binomial
- D) Exponencial

35. Completa el código para calcular una prueba t de Student en Python con SciPy:

```
from scipy import stats  
t_stat, p_value = stats.ttest_ind(grupo1, ____)
```

- A) `grupo2`
- B) `df`
- C) `group2.mean()`
- D) `p_value`

36. Detecta el error en este código de NumPy:

```
import numpy as np  
datos = [2, 4, 6, 8]  
desv = np.std{datos}
```

- A) `np.std` no existe
- B) `datos` debe ser un DataFrame
- C) Se usan `{ }` en lugar de `( )`
- D) `np` no puede calcular desviaciones

37. ¿Qué significa un error tipo I en pruebas de hipótesis?

- A) Aceptar  $H_0$  cuando es falsa

- B) Rechazar  $H_0$  cuando es verdadera
- C) Error en el cálculo del intervalo de confianza
- D) Error en la media muestral

38. Completa el código para obtener la varianza de una columna `ingresos` en Pandas:

```
var = df["ingresos"].____()
```

- A) `var()`
- B) `variance()`
- C) `std()`
- D) `desviacion()`

39. Ordena los pasos correctos para calcular una correlación entre dos columnas con Pandas:

- A) 1-2-3
- B) 2-1-3
- C) 3-1-2
- D) 1-3-2

40. La Ley de los Grandes Números establece que:

- A) La varianza disminuye a medida que aumenta la muestra
- B) La media muestral tiende a la media poblacional conforme aumenta el tamaño de muestra
- C) Los datos siempre convergen a distribución normal
- D) El error estándar aumenta con más datos

**Respuestas Módulo 4:** 31:B, 32:A, 33:B, 34:B, 35:A, 36:C, 37:B, 38:A, 39:A, 40:B.

## Módulo 5: Aprendizaje Supervisado (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em]¿Qué tipo de variable predice la regresión lineal?

41. A) Binaria
- B) Continua
- C) Categórica
- D) Texto

42. Completa el código para importar la clase `LinearRegression` desde scikit-learn:

```
from sklearn.linear_model import ____
```

- A) `LinearModel`
- B) `LinearRegressor`



- C) LinearRegression
- D) RegressionLinear

43. ¿Qué función activa usa la regresión logística?

- A) ReLU
- B) Sigmoid
- C) Tangente hiperbólica
- D) Softmax

44. Completa el código para entrenar un modelo de regresión logística en scikit-learn:

```
from sklearn.linear_model import LogisticRegression
model = LogisticRegression()
model.____(X_train, y_train)
```

- A) train
- B) fit
- C) predict
- D) transform

45. ¿Qué métrica es más adecuada para evaluar un modelo de clasificación binaria?

- A) RMSE
- B)  $R^2$
- C) Accuracy
- D) SSE

46. Completa el código para importar el algoritmo KNN en scikit-learn:

```
from sklearn.neighbors import ____
```

- A) KNeighborsClassifier
- B) KNNClassifier
- C) NearestNeighbors
- D) KNN

47. Detecta el error en este código de KNN:

```
from sklearn.neighbors import KNeighborsClassifier
knn = KNeighborsClassifier(k=5)
knn.fit(X_train, y_train)
```

- A) Falta neighbors=5, no k=5
- B) KNeighborsClassifier no existe
- C) fit no entrena en KNN

D) `y_train` no puede usarse con KNN

48. ¿Qué medida de impureza se utiliza en árboles de decisión?

A) Varianza explicada

B) Índice Gini

C)  $R^2$

D) Accuracy

49. Completa el código para importar un Random Forest de clasificación:

```
from sklearn.ensemble import ____
```

A) RandomForest

B) RandomForestClassifier

C) RFClassifier

D) DecisionForest

50. ¿Qué hace un clasificador SVM?

A) Calcula centroides para agrupar datos

B) Encuentra un hiperplano que maximiza el margen entre clases

C) Usa reglas if-else para separar datos

D) Calcula probabilidades de manera bayesiana

**Respuestas Módulo 5:** 41:B, 42:C, 43:B, 44:B, 45:C, 46:A, 47:A, 48:B, 49:B, 50:B.

## Módulo 6: Aprendizaje No Supervisado (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em]¿Cuál es el objetivo principal del algoritmo K-Means?

51. A) Maximizar la varianza entre variables

B) Minimizar la distancia entre puntos y sus centroides

C) Clasificar datos supervisados

D) Reducir dimensionalidad

52. Completa el código para importar KMeans desde scikit-learn:

```
from sklearn.cluster import ____
```

A) KMeans

B) KMeansClustering

C) ClusterK

D) KNN

53. En K-Means, ¿qué parámetro define el número de clusters?

- A) `max_iter`
- B) `random_state`
- C) `n_clusters`
- D) `k_neighbors`

54. Completa el código para entrenar un modelo K-Means con 3 clusters:

```
kmeans = KMeans(n_clusters=3)
kmeans.____(X)
```

- A) `train`
- B) `fit`
- C) `predict`
- D) `transform`

55. ¿Qué métrica interna se utiliza para evaluar la calidad de un clustering?

- A) Accuracy
- B) Inercia (SSE)
- C) RMSE
- D)  $R^2$

56. Completa el código para importar PCA desde scikit-learn:

```
from sklearn.decomposition import ____
```

- A) PCA
- B) `PrincipalComponentAnalysis`
- C) `PCAModel`
- D) `Decomposition`

57. ¿Qué hace PCA?

- A) Clasifica datos en clusters
- B) Reduce la dimensionalidad explicando la mayor varianza posible
- C) Calcula distancias entre observaciones
- D) Normaliza automáticamente los datos

58. Completa el código para aplicar PCA a 2 componentes principales:

```
pca = PCA(n_components=2)
X_pca = pca.____(X)
```

- A) `fit`
- B) `predict`
- C) `fit_transform`

59) ¿Qué ventaja tiene t-SNE frente a PCA?

- A) Es más rápido
- B) Captura relaciones no lineales en los datos
- C) Solo funciona con 3 dimensiones
- D) No requiere parámetros

60. Ordena los pasos correctos para aplicar t-SNE en scikit-learn:

- A) 2-1-3
- B) 1-2-3
- C) 3-2-1
- D) 1-3-2

**Respuestas Módulo 6:** 51:B, 52:A, 53:C, 54:B, 55:B, 56:A, 57:B, 58:C, 59:B, 60:B.

## Módulo 7: Ensemble, Bagging y Boosting (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em]¿Qué significa el método ensemble en aprendizaje automático?

61. A) Usar un único modelo optimizado  
B) Combinar varios modelos para mejorar el desempeño  
C) Reducir la dimensionalidad de los datos  
D) Usar únicamente modelos de regresión

62. Completa el código para importar un RandomForestClassifier:

```
from sklearn.ensemble import ____
```

- A) RandomForest
- B) RandomForestClassifier
- C) ForestClassifier
- D) BaggingClassifier

63. El concepto de bagging consiste en:

- A) Entrenar un solo modelo en todos los datos
- B) Entrenar múltiples modelos en subconjuntos de datos con reemplazo y promediar
- C) Usar pesos para dar más importancia a errores
- D) Usar modelos no supervisados para clasificación

64. Completa el código para crear un Random Forest con 200 árboles:

```
forest = RandomForestClassifier(n_estimators=____)
```

- A) 2

- B) 20
- C) 200
- D) 2000

65. Detecta el error en este código de Random Forest:

```
from sklearn.ensemble import RandomForestClassifier
model = RandomForestClassifier(n_estimators=100)
model.train(X_train, y_train)
```

- A) RandomForestClassifier no existe
- B) El método correcto es .fit() y no .train()
- C) n\_estimators no es un hiperparámetro válido
- D) X\_train no puede usarse en Random Forest

66. ¿Qué ventaja tiene un Random Forest frente a un árbol de decisión único?

- A) Siempre entrena más rápido
- B) Reduce el overfitting al promediar múltiples árboles
- C) No necesita datos de entrenamiento
- D) No requiere hiperparámetros

67. Completa el código para importar un GradientBoostingClassifier:

```
from sklearn.ensemble import ____
```

- A) GradientBoosting
- B) GradientBoostingClassifier
- C) GBMClassifier
- D) BoostingClassifier

68. En boosting, ¿cómo se entrenan los modelos?

- A) Todos en paralelo en subconjuntos de datos
- B) Secuencialmente, cada modelo corrige los errores del anterior
- C) Todos en el mismo conjunto y se promedian
- D) Únicamente con variables categóricas

69. Completa el código para crear un modelo AdaBoost en scikit-learn:

```
from sklearn.ensemble import AdaBoostClassifier
model = AdaBoostClassifier(n_estimators=50, ____=1.0)
```

- A) learning\_rate
- B) beta
- C) gamma

70. ¿Cuál de las siguientes es una desventaja de los métodos de boosting como XGBoost?

- A) Tienden a sobreajustar si no se regulan adecuadamente
- B) No pueden usarse en clasificación
- C) No funcionan con datos numéricos
- D) Siempre generan peor desempeño que un solo árbol

**Respuestas Módulo 7:** 61:B, 62:B, 63:B, 64:C, 65:B, 66:B, 67:B, 68:B, 69:A, 70:A.

## Módulo 8: Métricas de Desempeño (Regresión y Clasificación) (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em]¿Qué métrica de regresión calcula el Error Absoluto Medio?

71. A) RMSE  
B) MAE  
C)  $R^2$   
D) MSE

72. Completa el código para calcular el MSE en scikit-learn:

```
from sklearn.metrics import mean_squared_error
mse = mean_squared_error(y_true, ____)
```

- A) `y_train`
- B) `y_pred`
- C) `x_test`
- D) `x_train`

73. ¿Qué diferencia existe entre el MSE y el RMSE?

- A) El RMSE es la raíz cuadrada del MSE
- B) El MSE es la raíz cuadrada del RMSE
- C) El MSE solo se usa en clasificación
- D) El RMSE no depende del MSE

74. Completa el código para calcular el  $R^2$  en scikit-learn:

```
from sklearn.metrics import r2_score
r2 = r2_score(y_true, ____)
```

- A) `y_true`
- B) `y_pred`
- C) `x_train`
- D) `x_test`

75. ¿Qué significa un valor de  $R^2$  cercano a 1?

- A) El modelo tiene un mal ajuste
- B) El modelo explica casi toda la variabilidad de los datos
- C) Los errores son muy grandes
- D) Los datos no están correlacionados

76. Completa el código para importar la matriz de confusión:

```
from sklearn.metrics import ____
```

- A) `confusion_matrix_classification_report`
  - B) `accuracy_score_confusion`
77. ¿Qué representa la sensibilidad (recall) en un modelo de clasificación?
- A) La proporción de verdaderos negativos sobre todos los negativos
  - B) La proporción de verdaderos positivos sobre los positivos reales
  - C) La proporción de verdaderos positivos sobre los positivos predichos
  - D) La proporción de verdaderos negativos sobre los negativos predichos

78. Completa el código para calcular la precisión (precision):

```
from sklearn.metrics import precision_score  
prec = precision_score(y_true, y_pred, average="____")
```

- A) binary
  - B) multi
  - C) weighted
  - D) accuracy
79. ¿Qué curva se utiliza para evaluar el desempeño de un clasificador binario en distintos umbrales?
- A) Curva ROC
  - B) Histograma
  - C) Gráfico de dispersión
  - D) Boxplot
80. ¿Qué representa el AUC (Área Bajo la Curva ROC)?
- A) El área entre 0 y la curva de precisión
  - B) La capacidad del modelo para distinguir entre clases
  - C) El número de clusters encontrados
  - D) La varianza explicada

**Respuestas Módulo 8:** 71:B, 72:B, 73:A, 74:B, 75:B, 76:A, 77:B, 78:A, 79:A, 80:B.

## Módulo 9: Casos Prácticos de Código (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em] Detecta el error en este código para entrenar un modelo de regresión lineal:

81. 

```
from sklearn.linear_model import LinearRegression
model = LinearRegression
model.fit(X_train, y_train)
```

- A) `LinearRegression` no existe
- B) Falta instanciar: `LinearRegression()`
- C) `fit` no funciona en regresión
- D) `x_train` debe ser un `DataFrame`

82. Ordena los pasos para entrenar y evaluar un modelo en scikit-learn:

- A) 1-2-3-4-5
- B) 2-1-3-4-5
- C) 1-3-2-4-5
- D) 1-2-4-3-5

83. ¿Qué error hay en este código de PCA?

```
from sklearn.decomposition import PCA
pca = PCA(n_components=2)
X_pca = pca.fit(X_train, y_train)
```

- A) PCA no acepta `y_train` en `.fit()`
- B) `n_components=2` no es válido
- C) PCA no existe en sklearn
- D) `x_train` debe estar escalado con `RobustScaler`

84. Completa el código para dividir un dataset en entrenamiento y prueba:

```
from sklearn.model_selection import train_test_split
X_train, X_test, y_train, y_test = train_test_split(X, y, test_size=0.2, ____
=42)
```

- A) `random`
- B) `shuffle`
- C) `random_state`

85. Detecta el error en este código de KNN:

```
knn = KNeighborsClassifier(n_neighbors=5)
knn.fit(X_train)
```



- A) `KNeighborsClassifier` no existe
- B) Falta el argumento `y_train` en `.fit()`
- C) `n_neighbors=5` no es válido
- D) `X_train` debe estar normalizado

86. ¿Cuál es la mejor práctica antes de aplicar PCA?

- A) Normalizar los datos con `StandardScaler`
- B) Convertir todas las variables en categóricas
- C) Aumentar el número de dimensiones
- D) Entrenar un árbol de decisión

87. Completa el código para obtener un reporte de clasificación en sklearn:

```
from sklearn.metrics import classification_report
print(classification_report(y_true, ____))
```

- A) `y_true`
- B) `y_pred`
- C) `X_test`
- D) `X_train`

88. Ordena los pasos para aplicar un pipeline con escalado y regresión logística:

- A) 1-2-3-4
- B) 2-1-3-4
- C) 1-3-2-4
- D) 1-2-4-3

89. Detecta el error en este código de Random Forest:

```
forest = RandomForestClassifier()
forest.fit(X_train, y_train)
y_pred = forest.predict(X_test, y_test)
```

- A) `predict` no recibe `y_test` como argumento
- B) `RandomForestClassifier` no existe
- C) `fit` no funciona en Random Forest
- D) `X_test` debe estar escalado

90. ¿Qué librería de Python se recomienda usar para visualizaciones avanzadas de datos junto con Matplotlib?

- A) NumPy
- B) Pandas
- C) Seaborn
- D) scikit-learn

**Respuestas Módulo 9:** 81:B, 82:A, 83:A, 84:C, 85:B, 86:A, 87:B, 88:A, 89:A, 90:C.

## Módulo 10: Examen Integrador (15–20 min)

[label=0., leftmargin=\*, itemsep=0.6em]¿Qué algoritmo es más adecuado para predecir el precio de una casa?

- 91. A) Regresión lineal
- B) K-Means
- C) PCA
- D) t-SNE

92. Completa el código para entrenar un modelo de SVM lineal:

```
from sklearn.svm import SVC
svm = SVC(kernel="____")
svm.fit(X_train, y_train)
```

- A) linear
- B) rbf
- C) poly
- D) sigmoid

93. ¿Qué métrica es más adecuada para un problema de clasificación desbalanceada?

- A) Accuracy
- B) Recall
- C) RMSE
- D)  $R^2$

94. Completa el código para inicializar un KMeans con 4 clusters y semilla fija:

```
kmeans = KMeans(n_clusters=4, ____=42)
```

- A) seed
- B) `random_state`
- C) init

95. ¿Qué método ensemble entrena modelos en paralelo para luego promediar resultados?

- A) Boosting
- B) Bagging
- C) Stacking
- D) PCA

96. Detecta el error en este código de AdaBoost:

```
from sklearn.ensemble import AdaBoostClassifier
model = AdaBoostClassifier(n_estimators=50)
model.predict(X_train, y_train)
```

- A) `AdaBoostClassifier` no existe
- B) `predict` no recibe `y_train` como argumento
- C) `n_estimators` no es válido
- D) `fit` debe llamarse antes de `predict`

97. ¿Qué técnica de reducción de dimensionalidad es no lineal?

- A) PCA
- B) t-SNE
- C) Normalización Min-Max
- D) Bagging

98. Completa el código para calcular la matriz de confusión:

```
from sklearn.metrics import confusion_matrix  
cm = confusion_matrix(y_true, ____)
```

- A) `y_test`
- B) `y_pred`
- C) `x_train`
- D) `x_test`

99. ¿Qué ventaja tiene Gradient Boosting frente a Random Forest?

- A) Siempre entrena más rápido
- B) Se entrena secuencialmente corrigiendo errores previos
- C) No necesita parámetros de ajuste
- D) Siempre da menor error que cualquier otro modelo

100. Ordena los pasos correctos para aplicar PCA seguido de K-Means:

- A) 1-2-3-4
- B) 4-1-2-3
- C) 1-3-2-4
- D) 2-1-3-4

**Respuestas Módulo 10:** 91:A, 92:A, 93:B, 94:B, 95:B, 96:D, 97:B, 98:B, 99:B, 100:B.

---

*Nota:* Este examen combina preguntas teóricas y de código (completar, detectar errores, ordenar pasos). Use el espacio provisto para justificar si la certificación lo requiere.