

SAA: Test Capítulo 2

1. **Verdadero o Falso:** El objetivo principal del proyecto descrito en el documento es aprender sobre el negocio de bienes raíces.
 - a) Verdadero
 - b) Falso
2. ¿Cuál es el primer paso en un proyecto de aprendizaje automático según el documento?
 - a) Obtener los datos
 - b) Explorar los datos
 - c) Observar el panorama general
 - d) Seleccionar un modelo
3. ¿Qué tipo de tarea de aprendizaje automático es la predicción del precio medio de la vivienda en este proyecto?
 - a) Clasificación
 - b) Regresión
 - c) Agrupamiento
 - d) Refuerzo
4. **Verdadero o Falso:** El RMSE es una medida de rendimiento típica para problemas de clasificación.
 - a) Verdadero
 - b) Falso
5. ¿Qué significa RMSE?
 - a) Error Absoluto Medio
 - b) Error Cuadrático Medio Raíz
 - c) Desviación Absoluta Media
 - d) Error Cuadrático Medio
6. ¿Qué indica un coeficiente de correlación cercano a 1?
 - a) Una fuerte correlación negativa
 - b) Una fuerte correlación positiva
 - c) No hay correlación lineal
 - d) Una correlación no lineal
7. **Verdadero o Falso:** La función `scatter_matrix()` de Pandas se usa para visualizar la correlación entre atributos categóricos.
 - a) Verdadero
 - b) Falso
8. ¿Por qué es importante crear un conjunto de prueba al inicio del proyecto?
 - a) Para entrenar el modelo
 - b) Para visualizar los datos
 - c) Para evitar el sesgo de exploración de datos
 - d) Para mejorar el rendimiento del modelo

9. ¿Qué método de muestreo se recomienda cuando un atributo es muy importante para la predicción?
- a) Muestreo aleatorio simple
 - b) Muestreo estratificado
 - c) Muestreo por conglomerados
 - d) Muestreo sistemático
10. ¿Qué función se usa para dividir el conjunto de datos en subconjuntos de entrenamiento y prueba?
- a) `split()`
 - b) `shuffle_and_split_data()`
 - c) `train_test_split()`
 - d) `StratifiedShuffleSplit()`
11. **Verdadero o Falso:** El muestreo estratificado garantiza que cada subgrupo tenga una representación igualitaria en el conjunto de datos.
- a) Verdadero
 - b) Falso
12. ¿Cuál es el propósito de la función `value_counts()`?
- a) Calcular la media de los atributos numéricos
 - b) Mostrar un resumen estadístico de los atributos
 - c) Contar las ocurrencias de cada valor único en una columna
 - d) Visualizar la distribución de los datos
13. ¿Qué hace el método `info()` en un DataFrame de Pandas?
- a) Muestra las primeras filas del DataFrame
 - b) Proporciona información sobre el número de filas, tipos de datos y valores no nulos
 - c) Muestra un resumen estadístico de los atributos
 - d) Muestra la correlación entre los atributos
14. **Verdadero o Falso:** El método `describe()` muestra los valores nulos.
- a) Verdadero
 - b) Falso
15. ¿Qué hace la función `hist()` en un DataFrame de Pandas?
- a) Muestra las primeras filas del DataFrame
 - b) Proporciona información sobre el número de filas, tipos de datos y valores no nulos
 - c) Muestra un resumen estadístico de los atributos
 - d) Crea un histograma para cada atributo numérico
16. ¿Qué significa que un histograma esté "sesgado a la derecha"?
- a) Que los valores se concentran a la izquierda de la media
 - b) Que los valores se concentran alrededor de la media
 - c) Que los valores se extienden más hacia la derecha de la media que hacia la izquierda
 - d) Que la distribución de los datos es simétrica

17. ¿Cuál es el propósito de la clase **SimpleImputer**?

- a) Completar los valores faltantes
- b) Escalar los atributos numéricos
- c) Codificar atributos categóricos
- d) Seleccionar los atributos más importantes

18. **Verdadero o Falso:** **SimpleImputer** solo puede usar la mediana para completar los valores faltantes.

- a) Verdadero
- b) Falso

19. ¿Qué hace **OrdinalEncoder**?

- a) Crea un vector binario por categoría
- b) Convierte categorías de texto a números
- c) Escala los atributos numéricos
- d) Completa los valores faltantes

20. ¿Cuál es el resultado de aplicar **OneHotEncoder** a un atributo categórico?

- a) Un atributo numérico escalado
- b) Un vector binario por cada categoría
- c) Un atributo ordinal
- d) Un atributo con valores faltantes completados

21. **Verdadero o Falso:** **OneHotEncoder** crea una matriz dispersa por defecto.

- a) Verdadero
- b) Falso

22. ¿Cuál es la diferencia principal entre **MinMaxScaler** y **StandardScaler**?

- a) **MinMaxScaler** se usa para datos categóricos y **StandardScaler** para datos numéricos.
- b) **MinMaxScaler** centra los datos en cero y **StandardScaler** los escala entre 0 y 1.
- c) **MinMaxScaler** escala los datos a un rango específico, mientras que **StandardScaler** estandariza los datos.
- d) No hay diferencia entre ellos.

23. **Verdadero o Falso:** **MinMaxScaler** es más sensible a los valores atípicos que **StandardScaler**.

- a) Verdadero
- b) Falso

24. ¿Qué se entiende por "escalado de atributos"?

- a) Eliminar las filas que tengan valores faltantes.
- b) Transformar los valores de los atributos para que estén en la misma escala.
- c) Reducir la cantidad de atributos.
- d) Convertir los atributos numéricos a categóricos.

25. ¿Qué hace un **FunctionTransformer**?

- a) Completa valores faltantes
- b) Escala atributos
- c) Aplica una función a los datos
- d) Codifica atributos categóricos

26. **Verdadero o Falso:** Un `FunctionTransformer` no puede tener una función inversa.

- a) Verdadero
- b) Falso

27. ¿Qué clases se utilizan para crear un transformador personalizado en `scikit-learn`?

- a) `Estimator` y `Transformer`
- b) `BaseEstimator` y `PredictorMixin`
- c) `BaseEstimator` y `TransformerMixin`
- d) `Transformer` y `FunctionTransformer`

28. ¿Qué es un `pipeline` en `Scikit-Learn`?

- a) Una secuencia de pasos de transformación y un estimador final
- b) Un método para dividir datos en conjuntos de entrenamiento y prueba
- c) Un conjunto de hiperparámetros para un modelo
- d) Una función para visualizar datos

29. **Verdadero o Falso:** Un `pipeline` debe terminar con un transformador.

- a) Verdadero
- b) Falso

30. ¿Qué hace la clase `ColumnTransformer`?

- a) Completa valores faltantes en todo el dataset
- b) Escala todas las columnas a un rango específico
- c) Aplica diferentes transformaciones a diferentes columnas
- d) Combina todas las columnas en una sola

31. **Verdadero o Falso:** `ColumnTransformer` siempre devuelve una matriz dispersa.

- a) Verdadero
- b) Falso

32. ¿Qué hace la función `make_column_selector()`?

- a) Crea una matriz con las columnas seleccionadas
- b) Selecciona automáticamente las columnas de un tipo específico
- c) Elimina las columnas seleccionadas
- d) Renombra las columnas seleccionadas

33. ¿Qué clase se usa para realizar una búsqueda exhaustiva de hiperparámetros?

- a) `RandomizedSearchCV`
- b) `GridSearchCV`
- c) `HalvingGridSearchCV`
- d) `cross_val_score`

34. **Verdadero o Falso:** `GridSearchCV` evalúa todas las combinaciones posibles de hiperparámetros.

- a) Verdadero
- b) Falso

35.¿Qué ventaja tiene **RandomizedSearchCV** sobre **GridSearchCV**?

- a) **RandomizedSearchCV** es más preciso.
- b) **RandomizedSearchCV** es más eficiente cuando el espacio de hiperparámetros es grande.
- c) **RandomizedSearchCV** evalúa todas las combinaciones de hiperparámetros.
- d) **RandomizedSearchCV** siempre encuentra la mejor combinación de hiperparámetros.

36.¿Qué son los métodos de ensamble?

- a) Un tipo de transformador
- b) Un tipo de modelo que sólo procesa texto
- c) Modelos que combinan las predicciones de varios modelos para mejorar el rendimiento.
- d) Una técnica para dividir conjuntos de datos

37.**Verdadero o Falso:** La validación cruzada solo se usa para el modelo de arboles de decisión.

- a) Verdadero
- b) Falso

38.¿Qué función se usa para realizar validación cruzada?

- a) `train_test_split()`
- b) `GridSearchCV()`
- c) `cross_val_score()`
- d) `RandomizedSearchCV()`

39.¿Por qué es importante analizar los errores que comete un modelo?

- a) Para aumentar el rendimiento en el conjunto de prueba
- b) Para identificar posibles problemas con los datos o el modelo y buscar soluciones
- c) Para elegir el modelo más rápido
- d) Para saber qué hiperparámetros usar

40.¿Qué es la fuga de datos?

- a) Que los datos no tienen errores.
- b) Cuando un modelo es demasiado simple
- c) Que la información del conjunto de prueba se utiliza durante el entrenamiento
- d) Cuando un modelo hace un error.

41.**Verdadero o Falso:** Se debe evitar tocar el conjunto de prueba hasta el final del proyecto.

- a) Verdadero
- b) Falso

42.¿Qué se usa para guardar un modelo en **Scikit-Learn**?

- a) `pickle`
- b) `numpy`
- c) `joblib`
- d) `json`

43. ¿Cuál es el objetivo de monitorear un modelo en producción?

- a) Para mejorar la precisión del modelo en tiempo real.
- b) Para reducir los costos de computación
- c) Para detectar caídas en el rendimiento y errores en el sistema.
- d) Para cambiar los hiperparámetros del modelo

44. **Verdadero o Falso:** El rendimiento de un modelo en producción no necesita ser monitoreado si fue evaluado con la validación cruzada.

- a) Verdadero
- b) Falso

45. ¿Qué es el "model rot"?

- a) Un error en el código del modelo
- b) Una caída repentina del rendimiento
- c) El deterioro del rendimiento de un modelo con el tiempo debido a datos obsoletos.
- d) Una falla en la infraestructura

46. **Verdadero o Falso:** Los modelos no necesitan ser reentrenados si ya funcionan bien.

- a) Verdadero
- b) Falso

47. ¿Qué es ML Ops?

- a) Un nuevo algoritmo de Machine Learning.
- b) Una manera de compartir conjuntos de datos.
- c) Una disciplina que trata del ciclo de vida completo de un modelo de Machine Learning.
- d) Una biblioteca para gráficos.

48. **Verdadero o Falso:** La mayoría del trabajo en un proyecto de Machine Learning está en la selección de algoritmos avanzados.

- a) Verdadero
- b) Falso

49. ¿Qué se recomienda hacer si un atributo tiene una distribución de cola pesada?

- a) Eliminar el atributo
- b) Estandarizar el atributo
- c) Transformarlo aplicando el logaritmo o la raíz cuadrada.
- d) Escalar el atributo con **MinMaxScaler**.

50. **Verdadero o Falso:** Después de ajustar los hiperparámetros, el rendimiento en el conjunto de prueba debería ser siempre mejor que en la validación cruzada.

- a) Verdadero
- b) Falso

Soluciones

1.b	11.b	21.a	31.b	41.a
2.c	12.c	22.c	32.b	42.c
3.b	13.b	23.a	33.b	43.c
4.b	14.b	24.b	34.a	44.b
5.b	15.d	25.c	35.b	45.c
6.b	16.c	26.b	36.c	46.b
7.b	17.a	27.c	37.b	47.c
8.c	18.b	28.a	38.c	48.b
9.b	19.b	29.b	39.b	49.c
10.c	20.b	30.c	40.c	50.b