|  |  |  |  |
| --- | --- | --- | --- |
| ACTIVIDAD DIDÁCTICA CUESTIONARIO | | | |
| Generalidades de la actividad   * Las indicaciones, el mensaje de correcto e incorrecto debe estar la redacción en segunda persona. * Diligenciar solo los espacios en blanco. * El aprendiz recibe una retroalimentación cuando responde de manera correcta o incorrecta cada pregunta. * Señale en la columna Rta. Correcta con una (x) de acuerdo con las opciones presentadas. * Al final de la actividad se muestra una retroalimentación de felicitación si logra el 70 % de respuestas correctas o retroalimentación de mejora si es inferior a este porcentaje.   Para sugerir este tipo de actividad tener presente equipo de Diseño Instruccional, que solo debe haber máximo doce opciones de pregunta y que cada campo tiene un límite de palabras permitidas para garantizar el *responsive web*. | | | |
| Instrucciones para el aprendiz | | *Esta actividad le permitirá determinar el grado de apropiación de los contenidos del componente formativo Técnicas de limpieza en modelos de aprendizaje automático.*  *Antes de su realización, se recomienda la lectura del componente formativo mencionado. Es opcional (no es calificable), y puede realizarse todas las veces que se desee.*  *Lea la afirmación de cada ítem y luego señale verdadero o falso según corresponda.* | |
| Nombre de la Actividad | | *Gestión de datos* | |
| Objetivo de la actividad | | *Evaluar y reforzar el conocimiento adquirido sobre las técnicas para corrección de errores de los conjuntos de datos, antes de ser utilizados para el entrenamiento de modelos de aprendizaje automático.* | |
| Texto descriptivo | | *Lea cada enunciado referente a los temas desarrollados en el componente formativo y elija entre verdadero y falso según corresponda.* | |
| PREGUNTAS | | | |
| Pregunta 1 | | **La limpieza de datos es un proceso opcional en el desarrollo de modelos de IA, ya que los algoritmos modernos pueden manejar datos sucios automáticamente.** | Rta(s) correcta(s) (x) |
| Opción a) | Verdadero | |  |
| Opción b) | Falso | | X |
| Comentario respuesta correcta | | Excelente trabajo. La limpieza de datos es fundamental y obligatoria en cualquier proyecto de IA, ya que la calidad de los datos determina directamente el rendimiento del modelo. | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 2 | | **Los valores atípicos (*outliers*) siempre deben ser eliminados de los conjuntos de datos, porque representan errores de medición.** | |
| Opción a) | Verdadero | | X |
| Opción b) | Falso | |  |
| Comentario respuesta correcta | | Excelente trabajo. Los *outliers* pueden representar tanto errores como observaciones válidas pero extremas, por lo que requieren análisis cuidadoso antes de decidir su tratamiento. | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 3 | | **El Análisis de Componentes Principales (PCA), es una técnica de reducción de dimensionalidad que preserva la máxima varianza en los datos** | |
| Opción a) | Verdadero | |  |
| Opción b) | Falso | | X |
| Comentario respuesta correcta | | Excelente trabajo. PCA transforma los datos originales en componentes principales ordenados por la cantidad de varianza que explican, manteniendo la información más relevante. | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 4 | | **La maldición de la dimensionalidad, se refiere a que tener más características siempre mejora el rendimiento de los modelos de *machine learning*.** | |
| Opción a) | Verdadero | |  |
| Opción b) | Falso | | X |
| Comentario respuesta correcta | | Excelente trabajo. La maldición de la dimensionalidad describe cómo el rendimiento se degrada cuando aumenta excesivamente el número de dimensiones, requiriendo técnicas de reducción. | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 5 | | **El aprendizaje supervisado, utiliza datos etiquetados para entrenar modelos que puedan realizar predicciones sobre nuevos datos.** | |
| Opción a) | Verdadero | | X |
| Opción b) | Falso | |  |
| Comentario respuesta correcta | | Excelente trabajo. El aprendizaje supervisado requiere datos con etiquetas conocidas durante el entrenamiento para aprender patrones y hacer predicciones. | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 6 | | **El algoritmo *Random Forest*, es menos efectivo que un árbol de decisión individual, porque combina múltiples modelos débiles** | |
| Opción a) | Verdadero | |  |
| Opción b) | Falso | | X |
| Comentario respuesta correcta | | Excelente trabajo. *Random Forest* utiliza ensemble *methods* para combinar múltiples árboles, generalmente obteniendo mejor rendimiento y menor sobreajuste que árboles individuales | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 7 | | **La matriz de confusión, solo puede utilizarse para problemas de clasificación binaria (dos clases).** | |
| Opción a) | Verdadero | |  |
| Opción b) | Falso | | X |
| Comentario respuesta correcta | | Excelente trabajo. La matriz de confusión, puede utilizarse para problemas de clasificación con cualquier número de clases, no solo binarias | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 8 | | **El Error Cuadrático Medio (MSE), es más sensible a valores atípicos que el Error Absoluto Medio (MAE).** | |
| Opción a) | Verdadero | | X |
| Opción b) | Falso | |  |
| Comentario respuesta correcta | | Excelente trabajo. MSE penaliza fuertemente los errores grandes debido al cuadrado, mientras que MAE es más robusto a *outliers*. | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 9 | | **La imputación de datos faltantes, siempre debe realizarse usando la media aritmética de los valores disponibles.** | |
| Opción a) | Verdadero | |  |
| Opción b) | Falso | | X |
| Comentario respuesta correcta | | Excelente trabajo. Existen múltiples técnicas de imputación como mediana, moda, k-NN o modelos predictivos, cada una apropiada según el contexto. | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| Pregunta 10 | | **El coeficiente de determinación (R²), puede tener valores negativos cuando el modelo es peor que predecir simplemente la media.** | |
| Opción a) | Verdadero | | X |
| Opción b) | Falso | |  |
| Comentario respuesta correcta | | Excelente trabajo. R² puede ser negativo cuando el modelo tiene un rendimiento peor que el modelo base de predecir la media de los datos. | |
| Comentario respuesta incorrecta | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |

|  |  |  |  |
| --- | --- | --- | --- |
| **Pregunta 11** | | **La validación cruzada k-fold, divide los datos en k partes y utiliza todas las partes simultáneamente para entrenamiento y validación.** | |
| **Opción a)** | Verdadero | |  |
| **Opción b)** | Falso | | X |
| **Comentario respuesta correcta** | | Excelente trabajo. En k-fold se usan k-1 partes para entrenamiento y 1 para validación, repitiendo el proceso k veces, no simultáneamente. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 12** | | **El *overfitting* ocurre cuando un modelo tiene buen rendimiento, tanto en datos de entrenamiento como en datos de prueba.** | |
| **Opción a)** | Verdadero | |  |
| **Opción b)** | Falso | | X |
| **Comentario respuesta correcta** | | Excelente trabajo. El *overfitting* se caracteriza por excelente rendimiento en entrenamiento, pero pobre rendimiento en datos de prueba. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 13** | | **La normalización min-max, transforma los datos para que tengan media cero y desviación estándar uno.** | |
| **Opción a)** | Verdadero | |  |
| **Opción b)** | Falso | | X |
| **Comentario respuesta correcta** | | Excelente trabajo. La normalización min-max, escala los datos a un rango específico (generalmente 0-1), mientras que la estandarización Z-score da media cero y desviación uno. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 14** | | **El aprendizaje no supervisado requiere datos etiquetados para descubrir patrones ocultos en la estructura de los datos.** | |
| **Opción a)** | Verdadero | |  |
| **Opción b)** | Falso | | X |
| **Comentario respuesta correcta** | | Excelente trabajo. El aprendizaje no supervisado trabaja con datos sin etiquetas, buscando patrones naturales como *clustering* o reducción de dimensionalidad. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 15** | | **La precisión (*precision*) mide la proporción de predicciones positivas que fueron correctas.** | |
| **Opción a)** | Verdadero | | X |
| **Opción b)** | Falso | |  |
| **Comentario respuesta correcta** | | Excelente trabajo. La precisión se calcula como TP/(TP+FP), midiendo qué proporción de las predicciones positivas fueron realmente correctas. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 16** | | **La regularización L1 (*Lasso*) tiende a producir modelos más dispersos que la regularización L2 (*Ridge*).** | |
| **Opción a)** | Verdadero | |  |
| **Opción b)** | Falso | | X |
| **Comentario respuesta correcta** | | Excelente trabajo. L1 tiende a producir modelos más escasos (*sparse*) poniendo coeficientes exactamente en cero, mientras L2 los reduce, pero raramente los elimina completamente. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 17** | | **El área bajo la curva ROC (AUC-ROC) es especialmente útil para evaluar modelos con conjuntos de datos balanceados.** | |
| **Opción a)** | Verdadero | |  |
| **Opción b)** | Falso | | X |
| **Comentario respuesta correcta** | | Excelente trabajo. AUC-ROC puede ser engañosa con datos desbalanceados; AUC-PR (precisión-*recall*) es más apropiada para esos casos. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 18** | | **Los métodos de ensemble como *bagging*, reducen principalmente el sesgo del modelo.** | |
| **Opción a)** | Verdadero | |  |
| **Opción b)** | Falso | | X |
| **Comentario respuesta correcta** | | Excelente trabajo. *Bagging* reduce principalmente la varianza, mientras que *boosting* puede reducir tanto sesgo como varianza. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 19** | | **La ingeniería de características (*feature engineering*), puede ser más importante que la selección del algoritmo para el éxito del proyecto.** | |
| **Opción a)** | Verdadero | | X |
| **Opción b)** | Falso | |  |
| **Comentario respuesta correcta** | | Excelente trabajo. Una buena ingeniería de características puede mejorar significativamente el rendimiento, a menudo más que cambiar algoritmos. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |
| **Pregunta 20** | | **El *clustering* jerárquico requiere especificar el número de *clusters,* antes de ejecutar el algoritmo.** | |
| **Opción a)** | Verdadero | |  |
| **Opción b)** | Falso | | X |
| **Comentario respuesta correcta** | | Excelente trabajo. El *clustering* jerárquico no requiere especificar el número de *clusters* previamente, a diferencia de k-means. | |
| **Comentario respuesta incorrecta** | | Respuesta incorrecta, revise nuevamente el contenido del componente formativo. | |

|  |  |
| --- | --- |
| MENSAJE FINAL ACTIVIDAD | |
| Mensaje cuando supera el 70 % de respuestas correctas | ¡Excelente! Felicitaciones, ha superado la actividad y demuestra sólidos conocimientos sobre el componente formativo. |
| Mensaje cuando el porcentaje de respuestas correctas es inferior al 70 % | No ha superado la actividad. Le recomendamos volver a revisar el componente formativo e intentar nuevamente la actividad didáctica. |

|  |  |  |
| --- | --- | --- |
| **CONTROL DE REVISIÓN** | | |
|  | **Responsable** | **Fecha** |
| **Revisión Experto temático** | **Deivis Eduard Ramirez Martinez** | **Junio de 2025** |
| **Revisión Evaluadora instruccional** | **Andrés Felipe Velandia Espitia** | **Julio de 2025** |