

IMPLICACIONES TÉCNICAS Y PRÁCTICAS DE LAS REDES ADVERSARIAS GENERATIVAS A LA CIENCIA ABIERTA EN EDUCACIÓN



Bethencourt-Aguilar, A., Castellano-Nieves, D., Sosa-Alonso, J.J. y Area-Moreira, M. (2022). Implicaciones técnicas y prácticas de las Redes Adversarias Generativas a la Ciencia Abierta en Educación. *RiITE Revista interuniversitaria de investigación en Tecnología Educativa*, 13, 138-156. <https://doi.org/10.6018/riite.545881>

¿Cómo afrontar el problema del tamaño pequeño de la muestra?



Gracias al desarrollo de la IA:

REDES GENERATIVAS ADVERSARIAS (GAN)

Es un tipo de modelo generativo que permite generar muestras sintéticas con la misma distribución que los datos reales

Es un método muy eficaz pero no se ha estudiado su aplicabilidad en el campo educativo

CIENCIA ABIERTA

Para que se favorezca el intercambio de información y colaboración entre los agentes educativos y sociedad

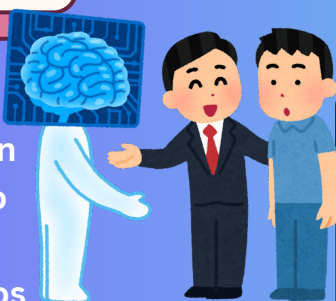
- Transparencia
- Reproducibilidad
- Intercambio
- Disposición pública



Difusión sin atender a los derechos de protección de los datos

MÉTODO

Primer acercamiento a la utilización de los datos sintéticos en el ámbito de la investigación educativa y las aportaciones que la ciencia de datos puede ofrecer a la Ciencia Abierta en el ámbito educativo



Procedimiento de generación del conjunto sintético

Filtrado del conjunto de datos original.

Creación del conjunto de datos artificial:

- Preprocesamiento de los datos (extracción de características, imputación, escalado, GANs)
- Creación de los modelos
- Entrenamiento de los modelos
- Evaluación de los modelos
- Generación de datos sintéticos
- Evaluación de los datos generados

Objetiva

Generar y validar un conjunto de datos sintéticos entrenados sobre Tecnología Educativa partiendo de un instrumento de encuesta que permita la replicabilidad y la sustitución del conjunto de datos original, detectando las complejidades y los principales resultados destacables del proceso

Participantes

El conjunto de datos originales se obtuvo de una encuesta al alumnado de postgrado de la Universidad de La Laguna en el año académico 2021

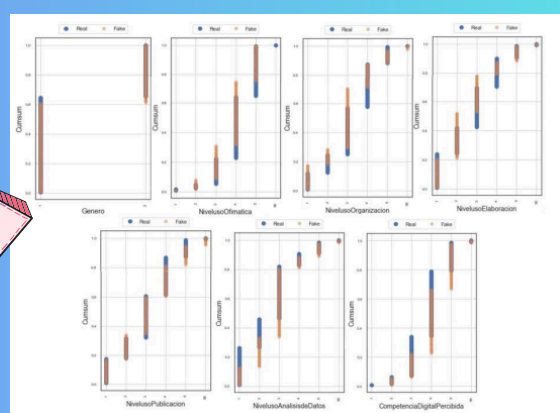
RESULTADOS

1. Representación visual de variables en comparativa con los datos reales y los datos sintéticos
2. Validación estadística
3. Análisis

Los conjuntos de datos sintéticos no mantenían la estructura de datos de los sujetos característica del conjunto original

4. Nuevos procedimientos hacia las CopulaGAN-WAN ajustando las correlaciones

Los resultados muestran equivalencia en grado de similitud de los datos originales con los datos creados



	Nivel uso Ofimática	Nivel uso Organización	Nivel uso Elaboración	Nivel uso Publicación	Nivel uso Búsqueda	Nivel uso Análisis de Datos
Nivel uso Ofimática	1	,463**	,183**	,158*	,481**	,166*
Nivel uso Organización	,463**	1	,359**	,374**	,457**	,213**
Nivel uso Elaboración	,183**	,359**	1	,471**	,201**	0,126
Nivel uso Publicación	,158*	,374**	,471**	1	,248**	,186**
Nivel uso Búsqueda	,481**	,457**	,201**	,248**	1	,294**
Nivel uso Análisis de Datos	,166*	,213**	0,126	,186**	,294**	1

CONCLUSIONES

Las GAN no solo pueden aportar a las investigaciones en TE permitiendo mayor difusión de información, recursos digitales y conjuntos de datos sino que además el procedimiento está estrechamente ligado con la TE porque incluye el uso de algoritmos con fines educativos o académicos y forma parte del ámbito con la mayor cercanía a la tecnología digital, ingeniería de datos e Inteligencia Artificial.



New Tab

PROBLEMÁTICAS

- Creación de las tablas donde se mantenga la relación entre las variables
- Creación de otros datos sintéticos que incluya la misma estructura de datos original
- Limitaciones de calidad de la muestra original
- Uso éticamente responsable

VENTAJAS

- Creación de datos sintéticos que incrementan el número de registros muestrales disponibles con un coste mínimo
- Anonimización de datos facilita la ciencia abierta