



**Implicaciones técnicas y
prácticas de las Redes
Adversarias Generativas a la
Ciencia Abierta en
Educación**

**Technical and Practical Implications of
Generative Adversarial
Networks for Open Science in
Education**

Anabel Bethencourt-Aguilar



Universidad de la Laguna (España)


abethenc@ull.edu.es

Dagoberto Castellano-Nieves 

Universidad de la Laguna (España)
dcastell@ull.edu.es

Juan José Sosa-Alonso 

Universidad de la Laguna (España)
jsosalo@ull.edu.es

Manuel Area-Moreira 

Universidad de la Laguna (España)
manarea@ull.edu.es

Recibido: 05/11/2022 Aceptado: 1/12/2022 Publicado:
1/12/2022

RESUMEN

Las Redes Adversarias Generativas (GAN), propias de la Inteligencia Artificial, permiten la creación de datos sintéticos anonimizados útiles para hacer Ciencia Abierta dentro de la investigación educativa. El presente estudio realiza una experimentación en la creación de datos artificiales a partir de un conjunto de datos obtenidos de una encuesta sobre niveles de uso de herramientas

digitales y la frecuencia de actividades personales con tecnología. Los datos originales pertenecen a una muestra de alumnado de las titulaciones de postgrado de la Universidad de La Laguna. Los resultados muestran un grado de similitud adecuado entre el conjunto de datos original y el conjunto creado artificialmente a través de algoritmos predictivos. La obtención de conjuntos de datos sintéticos equivalentes a los originales en estructura, forma y extensión permite la liberación de los datos a la comunidad académica salvaguardando la protección de la información confidencial y contrastando una técnica que permite impulsar la Ciencia Abierta desde la obtención y tratamiento de los datos. Las Redes Adversarias Generativas pueden ser utilizadas en la investigación educativa con fines hacia la transparencia en los procedimientos metodológicos y técnicos y a la difusión de conjuntos de datos para fines académicos, investigativos y educativos.

PALABRAS CLAVE

Inteligencia artificial; datos sintéticos; protección de la privacidad; Ciencia Abierta; Tecnología Educativa

~ 138 ~

RiiTE, Núm. 13 (2022), 138-156

Implicaciones técnicas y prácticas de las Redes

ABSTRACT

Generative Adversarial Networks (GANs), which are characteristic of Artificial Intelligence, allow the creation of synthetic anonymised data useful for Open Science in educational research. This study experiments with the creation of artificial data from a dataset obtained from a survey on levels of use of digital tools and frequency of personal activities with technology. The original data belong to a sample of students from postgraduate degrees at the University of La Laguna. The results show an adequate degree of similarity between the original data set and the set artificially created through predictive algorithms. Obtaining synthetic datasets equivalent to the original ones in structure, shape and extension allows the release of the data to the academic community, safeguarding the protection of confidential information and contrasting a technique that allows the promotion of Open Science from the collection and processing of the data. Generative Adversarial Networks can be used in educational research for the purpose of transparency in methodological and technical procedures and the dissemination of datasets for academic, research and

educational purposes.

KEYWORDS

Artificial Intelligence; Synthetic Data; Privacy Protection; Open Science; Educational Technology

CITA RECOMENDADA:

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>

Principales aportaciones del artículo y futuras líneas de investigación:

- Análisis de las ventajas de las Redes Adversarias Generativas para la Ciencia Abierta
- Comprobación de la técnica y del grado de similitud de los conjuntos originales y sintéticos
- Indicación de las principales consideraciones para la generación de conjuntos sintéticos
- Líneas futuras en la investigación de datos sintéticos

1. INTRODUCCIÓN

1.1. Las Redes Adversarias Generativas en educación

El aprendizaje profundo, desarrollado por Hinton y Salakhutdinov (2006), se ha convertido en la tecnología clave de la inteligencia de los grandes datos (Zhuang et al., 2017) y ha dado lugar a

importantes avances científicos en la temática (Al-Qizwini et al., 2017; Wang & Sng, 2015; Mohamed et al., 2012; Jones, 2015).

A diferencia de las metodologías clásicas de aprendizaje automático estadístico, el aprendizaje profundo, como núcleo del método de inteligencia de grandes datos, tiene una estructura de modelo relativamente compleja que depende del tamaño y la calidad de un conjunto de datos (Goodfellow et al., 2014). Se puede afirmar que el rendimiento de un modelo de aprendizaje profundo viene determinado por el tamaño y la calidad de los datos con los que se trabaja.

Sin embargo, en muchas aplicaciones del mundo real resulta difícil y costoso obtener datos de muestra “etiquetados”, con niveles y valores de las variables identificados. Además del alto coste económico, las características de los datos son complicadas y tienen una alta dimensionalidad,

dando lugar a lo que se conoce por “el problema del tamaño pequeño de la muestra” (Theodoridis & Koutroumbas, 2006) que limita las posibilidades de generalización en los procesos de aprendizaje. Para afrontar el problema de la escasez de datos se han propuesto varios métodos de sobremuestreo, cuya principal ventaja es que son autosuficientes (DeRouin & Brown, 1991). Uno de los más conocidos es el sobremuestreo de minorías sintéticas (SMOTE) (Chawla et al., 2002). A partir del SMOTE se han desarrollado varias alternativas (Han et al., 2005; He et al., 2008; Barua et al., 2014; Xie et al., 2015; Douzas & Bacao, 2017), la mayoría de las cuales se centran en el aprendizaje por desequilibrio, en el que se puede conseguir un mejor rendimiento añadiendo instancias de sobremuestreo al conjunto de datos de la clase minoritaria. Sin embargo, los conjuntos de datos de muchos ámbitos siguen siendo insuficientes, más que desequilibrados, en cada clase.

Otros enfoques tradicionales, distintos del SMOTE, también dirigidos a abordar el problema tamaño de

la muestra han sido el Bishop (1995) o el de Zhou y Jiang (2004) que entrena primero una red neuronal y luego la emplea para generar un nuevo conjunto de entrenamiento. Li y Lin (2006) idean un método de generación de muestras virtuales basado en una estimación de la densidad del núcleo internalizada y, también, Li y Fang (2009) propusieron de nuevo una técnica de generación de muestras virtuales no lineales utilizando el descubrimiento de grupos y las ecuaciones paramétricas de la hiperesfera. Sin embargo, estos métodos no pueden hacer uso de las características inherentes a las muestras, lo que supone una limitación de los modelos de entrenamiento.

Como alternativa, a raíz del rápido desarrollo de la nueva generación de inteligencia artificial (IA), ha surgido la posibilidad de abordar este problema haciendo uso de Redes Generativas Adversativas (GAN) basadas en una red neuronal profunda (DNN) que consiste en dos redes neuronales que compiten mutuamente entre sí y que se implementan a través de algoritmos de IA para el aprendizaje no supervisado. Con la ventaja añadida

de que esta estrategia permite la aplicación del aprendizaje profundo en el caso de un tamaño de muestra pequeño.

Una GAN (Goodfellow et al., 2014) es un potente tipo de modelo generativo (Wang et al., 2017), que puede utilizarse para generar muestras sintéticas con la misma distribución que los datos reales para resolver el problema de la insuficiencia de datos disponibles (Creswell et al., 2018). La arquitectura profunda de una GAN consta de dos funciones (generador y discriminador), que pueden “aprender” simultáneamente de los datos de entrenamiento de forma adversarial (Radford et al., 2015). En el proceso de aprendizaje, el generador capta la distribución potencial de los datos reales y genera muestras sintéticas, mientras que el algoritmo discriminador intenta diferenciar entre las muestras reales y las sintéticas con la mayor precisión posible.

Se ha demostrado que un GAN puede aplicarse con éxito a la generación de imágenes (Santana & Hotz, 2016; Gou et al., 2016), al procesamiento del lenguaje (Li et al., 2017; Pascual et al., 2017) y al aprendizaje supervisado con datos de entrenamiento

insuficientes. Estos estudios demuestran que las muestras sintéticas generadas por un GAN se ajustan a la distribución de las muestras originales. Además, el éxito de las GAN en varios campos indica que este modelo generativo es independiente del conocimiento preciso del dominio o constructo teórico subyacente, lo que es propicio para la aplicación de este enfoque en otros campos. Desde la perspectiva del aprendizaje supervisado con datos de entrenamiento insuficientes, la mayoría de

~ 140 ~

RiiTE, Núm. 13 (2022), 138-156

Implicaciones técnicas y prácticas de las Redes
Adversarias

Generativas a la Ciencia Abierta en Educación

los estudios existentes tratan el problema del desequilibrio de clases, verificándose que la solución por la vía GAN es el método más eficaz para incrementar el tamaño de la muestra (Fiore et al., 2019; Douzas & Bacao, 2018).

La ampliación de la escala de datos mediante el aumento de los mismos permite mejorar el

rendimiento del aprendizaje automático supervisado basado en las GAN en diversos campos. Sin embargo, a pesar del notable éxito de las GAN, hasta donde sabemos, aún no se ha estudiado su aplicabilidad con un tamaño de muestra pequeño en el campo educativo, donde típicamente se trabajan con muestras con datos escasos.

1.2. Implicaciones hacia la Ciencia Abierta

La UNESCO (2021) define la “Ciencia Abierta” como “un constructo inclusivo que combina diversos movimientos y prácticas con el fin de que los conocimientos científicos multilingües estén abiertamente disponibles y sean accesibles para todos, así como reutilizables por todos” (UNESCO, 2021). De este modo, la UNESCO defiende este nuevo paradigma hacia las nuevas tendencias académicas, investigadoras y docentes en el que se favorezca el intercambio de información y colaboración entre los agentes educativos y sociedad, de común acuerdo para el beneficio del conjunto de la población y se permita la creación de

nuevos procesos y dinámicas en esta etapa de transformación del concepto de ciencia. La Ciencia Abierta se convierte en uno de los principios actuales que promueven las organizaciones junto con la educación inclusiva, de calidad, solidaria e innovadora (González-Pérez et al., 2022; Ramírez-Montoya et al., 2022).

En la literatura académica especializada en Ciencia Abierta se subraya las bondades que tiene este cambio de planteamiento para el incentivo de la transparencia, reproducibilidad y la realización de nuevos estudios a raíz de la colaboración, el intercambio de datos y la disposición pública (Abadal & Anglada, 2020; Strcic et al., 2022; Alés, 2020). El fenómeno a favor de la Ciencia Abierta se ensambla a través del impulso en transparencia en el marco de la investigación, tecnología e innovación y tiene su especial presencia en las universidades como representantes de espacios hacia la generación y transferencia de conocimientos (Alés, 2020). Dentro de la amplitud del constructo también Ciencia Abierta supone la disposición del público del conjunto de los datos sin procesar con vías de

facilitar la transparencia a las revistas académicas de los análisis y resultados extraídos concibiendo que estas nuevas prácticas pueden tener su impacto en todas las etapas del “hacer ciencia” (Strcic et al., 2022; Abadal & Anglada, 2020; Van der Zee & Reich, 2018).

La sociedad actual demanda grandes cantidades de datos con los que poder responder a las necesidades actuales con bases informadas y coherentes a las complejidades presentes. Este cambio de paradigma incluye una modificación en el modo en el que nos relacionamos con la información y el contenido y en la forma en la que ésta adquiere importancia para nuestra realidad. A modo de ejemplo, tras la situación vivenciada por la COVID-19 la difusión de información, datos, ensayos y librerías fue y sigue siendo de vital importancia para la rápida actuación con bases científicas y evidenciados en datos que garanticen la mejor respuesta ante una situación como la vivenciada (Strcic et al., 2022). La puesta en marcha de proyectos de investigación en la época actual también supone el conocimiento

de cómo gestionar la información y las posibilidades del intercambio de datos para beneficio de la comunidad académica e incluye, por lo tanto, el encuentro con múltiples complicaciones y dificultades que

~ 141 ~

Anabel Bethencourt, Dagoberto Castellanos
RiiTE, Núm. 13 (2022), 138-156 Juan José Sosa y Manuel
Area

implican, la realización de Ciencia Abierta en todas sus fases, un desafío (Fressoli & Arza, 2018). Del mismo modo, en la actualidad empieza a formar parte de algunos requisitos de agencias de investigación, financiación o requisito de revistas y comités evaluadores precisamente la disposición pública de los datos (Logan et al., 2021; Van Dijk et al., 2021) y, por lo tanto, se presta cada vez más relevante las investigaciones con equipo multidisciplinares en los que puedan colaborar hacia procesos más éticos y transparentes sin que los conocimientos técnicos sean en sí mismo el mayor

problema (Parti & Szigeti, 2021).

Las tecnologías digitales forman parte indispensable de este nuevo replanteamiento de los modos en los que investigamos, compartimos, difundimos y comunicamos (Van der Zee & Reich, 2018; Burgos, 2020; Conole & Brown, 2018; González-Pérez et al., 2022). El movimiento de la Ciencia Abierta debe aprovechar las prácticas y tecnologías digitales que permita la realización de su cometido y la puesta a disposición de los datos e informaciones sobre el procedimiento conceptual, metodológico, técnico y práctico de lo que acontece a nuestros estudios emergentes. Desde el otro prisma, la Ciencia Abierta puede favorecer a que, los principios y planteamientos de la tecnología educativa adquieran mayor sustento a través del conocimiento de los datos y realizar toma de decisiones más ajustados al contexto en el que se enmarca (Alhadad et al., 2018; Bethencourt-Aguilar et al., 2021ayb).

Este cambio de paradigma y de perspectiva cultural fuerza a la búsqueda de nuevas estrategias y técnicas que permitan eliminar la información confidencial entre los agentes que intervengan en nuestras

investigaciones e impulsen nuevos procedimientos de anonimización en los datos que reduzcan las posibilidades de identificación o reidentificación de la población o muestra de estudio y que permita la difusión de dicha información entre la comunidad. En este sentido es donde la Inteligencia Artificial, y en concreto, las Redes Adversarias Generativas, pueden impulsar estos procedimientos hacia la difusión de los conjuntos de datos con fines académicos e investigadores sin atender a los derechos de protección y permitir así la liberación de datos en nuestro ámbito educativo que impulse la puesta a prueba de novedosas técnicas de análisis, realizar nuevas interpretaciones a los datos o facilitar la disposición de material con el que nuestro alumnado pueda aprender.

2. MÉTODO

Este trabajo es un 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. La creación de un conjunto de datos generados a través de Inteligencia Artificial que mantenga una adecuada similitud a la base de datos original y que pueda sustituir al conjunto de datos original es relevante para la difusión y compartición de datos de carácter educativo salvaguardando la información confidencial de los agentes implicados y anonimizando los datos para su posterior utilización.

2.1. Objetivos de investigación

El objetivo de esta investigación, por lo tanto, es 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.

~ 142 ~

RiiTE, Núm. 13 (2022), 138-156

Implicaciones técnicas y prácticas de las Redes
Adversarias

2.2. Participantes

El conjunto de datos originales se ha obtenido tras la aplicación de una encuesta dirigida al alumnado de titulaciones de postgrado de la Universidad de La Laguna en el año académico 2021 posterior al impacto del COVID en la docencia universitaria. La encuesta obtiene un total de 240 registros y 174 variables de interés sobre las percepciones, necesidades y dificultades en relación con la tecnología educativa, las titulaciones académicas cursadas, los planteamientos didácticos metodológicos de las aulas virtuales y el impacto de la COVID-19 en el marco de las asignaturas.

Las variables que sometemos a prueba en este estudio en concreto han sido las siguientes: **Tabla 1.**

Puntuaciones del clasificador y similitudes Jaccard

Categorías	Variables
Nivel de uso de herramientas	Herramientas de ofimática

digitales Herramienta para la
organización, productividad y colaboración
Herramientas para la elaboración y creación de contenidos
multimedia
Herramientas para la publicación y difusión
Herramientas para la búsqueda y recuperación de
información
Herramientas para el análisis de datos e investigación
Actividades Frecuencia de uso de las
redes sociales
personales con TIC
Frecuencia de navegar por Internet
Frecuencia de escucha de la radio
Frecuencia de visualización de la televisión
Frecuencia de visualización de series o películas
Frecuencia de uso de videojuegos
Frecuencia de visualización de vídeos en plataformas
Youtube
de difusión como Frecuencia de seguir conferencias,
charlas, webinar en directo o
previamente grabado
Frecuencia de comprar tecnología, ropa u otros, vía online
Frecuencia de realizar videollamadas con amigos/as o
familiares a través de
la tecnología
Frecuencia de lectura en ebook de literatura

La muestra total obtenida con 240 registros del alumnado de las titulaciones de postgrado incluye una mayor frecuencia de mujeres ($n=162$) que hombres ($n=78$). La mayoría del alumnado se encuentra entre la franja de edad de entre los 21 y 25 ($n=113$), con menor frecuencia conforme

~ 143 ~

Anabel Bethencourt, Dagoberto Castellanos

RiiTE, Núm. 13 (2022), 138-156 Juan José Sosa y Manuel

Area

se aumenta el rango de edad como entre 26 y 30 ($n=59$) y entre 31 y 35 años ($n=39$). En cuanto a la rama de conocimiento del que procede, la mayoría del alumnado pertenece a titulaciones de postgrado de Ciencias Sociales y Jurídicas ($n=80$), Ciencias ($n=51$), Artes y Humanidades ($n=45$), Ciencias de la Salud ($n=41$) y, por último, con menor nivel de frecuencia en la muestra de nuestro conjunto de datos, Ingeniería y Arquitectura ($n=23$).

2.3. Procedimiento de generación del conjunto sintético

Los primeros pasos en el procedimiento fue el filtrado del conjunto de datos original, eliminando aquellos registros incompletos o nulos que se encontraban en nuestra base de datos. El conjunto de datos final del que parte el siguiente análisis lo constituye un total de 127 registros de alumnado y 34 variables con diseño categórico ordinal en su mayoría.

El procedimiento se puede dividir en dos partes fundamentales, por un lado, la generación de los datos sintéticos y posterior la evaluación de la calidad del modelo obtenido. La experimentación se realizó con la biblioteca TGAN y se evalúa a través de diferentes métricas que inducen a la verificación de la calidad del modelo. La calidad del modelo resultante se evalúa desde el grado de similitud de la generación del conjunto de tablas y en el grado de validez en el que la biblioteca utilizada para la

creación de los datos sintéticos muestra a través de sus estadísticos. En este proceso en dos partes de generación y validación, las métricas para la validación incluyen el análisis en las propiedades estadísticas, la eficacia del aprendizaje automático y la privacidad, consecuencia de los algoritmos predictivos de regresión y clasificación en nuestras variables. Las estrategias predictivas practican la generación y validación del conjunto de datos con una técnica cruzada originada del entrenamiento con una parte de los datos y la evaluación con el otro.

El procedimiento, por lo tanto, realizado para la creación del conjunto de datos artificial muestra los pasos habituales en este tipo de estudios: (a) Preprocesamiento de los datos (extracción de características, imputación, escalado, GANs); (b) Creación de los modelos; (c) Entrenamiento de los modelos; (d) Evaluación de los modelos; (g) Generación de datos sintéticos; (h) Evaluación de los datos generados.

3. RESULTADOS

El propio proceso de creación de datos sintéticos incluye en su diseño, un grado ajustado y coherente de similitud a la tabla de datos original. En nuestro ejemplo de tabla de datos, se muestra un nivel bastante alto de grado de similitud de las distintas bases de datos sintéticas solicitadas en relación con los datos originales.

En el siguiente gráfico se puede apreciar la representación visual de algunas de nuestras variables en comparativa con los datos reales y los datos sintéticos. Los datos sintéticos imitan las representaciones obtenidas en el conjunto original y replican a un nuevo conjunto de datos.

~ 144 ~

RiiTE, Núm. 13 (2022), 138-156

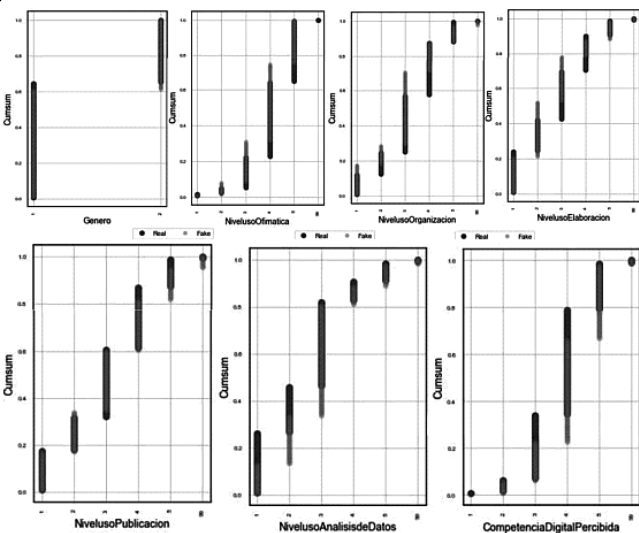
Implicaciones técnicas y prácticas de las Redes Adversarias

Generativas a la Ciencia Abierta en Educación

Figura 1.

Representación de datos originales y sintéticos en las variables de uso de herramientas digital



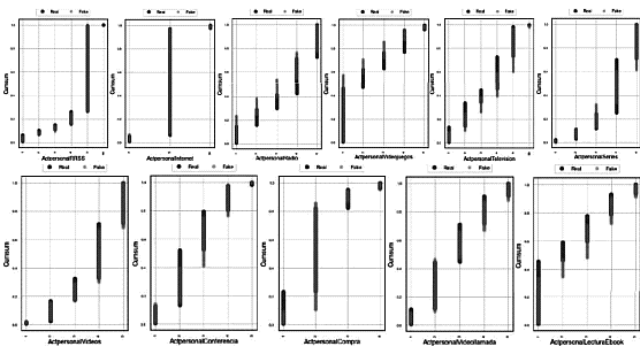


Estas gráficas representan la distribución de los datos reales y su distribución en los sintéticos en cada una de las variables y en sus distintas opciones de respuestas. De este modo, se puede comparar de forma gráfica lo ajustados que están los datos sintéticos con los reales en cada una de las variables consideradas en nuestro trabajo. El mismo resultado

se obtiene en el conjunto de variables sobre actividades personales del alumnado que incluya el uso de la tecnología digital (Figura 2).

Figura 2.

Representación de datos originales y sintéticos en las variables de uso de actividades personales con TIC



~ 145 ~

Anabel Bethencourt, Dagoberto Castellanos
RiiTE, Núm. 13 (2022), 138-156 Juan José Sosa y Manuel Area

Estas representaciones muestran lo ajustados de los datos generados, y aunque en la propia

generación de los datos se asegure la similitud de ambos conjuntos, es importante igualmente la validación estadística que evidencie del mismo modo lo similares que pueden llevar a ser el conjunto de datos sintéticos con el original. Los estadísticos ejecutados en la base de datos real y sintética son generados con la librería utilizada y permiten acercarse con datos reales si los dos conjuntos (original y sintético) pueden ser similares. En otras palabras, las siguientes métricas nos permiten valorar qué grado de validez tiene el conjunto sintético y el original en conjunto y por ello si un modelo podría ser intercambiable y sustituible por el original sin perder las características estadísticas de los datos originales.

Tabla 2.
Puntuaciones del clasificador y similitudes Jaccard

	f1_real	f1_fake	jaccard_similarity
LogisticRegression_real_testset			0.2308
LogisticRegression_fake_testset			0.4615
RandomForestClassifier_real_testset			0.4231
RandomForestClassifier_fake_testset			0.1923

DecisionTreeClassifier_real_testset	0.3077
DecisionTreeClassifier_fake_testset	0.3077
MLPClassifier_real_testset	0.3846
MLPClassifier_fake_testset	0.3462

Tabla 3.

Resultados variados y puntuación de similitud en conjunto sintético y original

0.3846	0.2683
0.4615	0.3333
0.4231	0.3333
0.2692	0.2381
0.4615	0.1304
0.3462	0.1818
0.5000	0.3000

0.3462

0.2381

Result	Result
Column Correlation Distance RMSE	0.0668
Column Correlation distance MAE	0.0347
Duplicate rows between sets (real/fake)	(0, 0)
nearest neighbor mean	5.1070
nearest neighbor std	0.2976
Basic statistics	1.0000
Correlation column correlations	0.7715
Mean Correlation between fake and real columns	0.72601
MAPE Estimator results	0.7510
Similarity Score	0.8121

En la siguiente tabla, se evidencia que distintas solicitudes de creación de conjuntos de datos pueden mostrar distintos índices de calidad de similitud del conjunto sintético con el conjunto comparativo original (Tabla 4). El grado de similitud, por lo tanto, de la tabla de datos sintética podría variar del proceso de entrenamiento y del conjunto de datos extraído. Existe variabilidad

~ 146 ~

RiiTE, Núm. 13 (2022), 138-156

Implicaciones técnicas y prácticas de las Redes Adversarias

Generativas a la Ciencia Abierta en Educación

en el índice de puntuación de similitud, pero todas tienen un nivel medio adecuado que favorecen la utilización de la tabla de datos sintética en sustitución a la original.

Tabla 4.

Resultados entre distintos órdenes de conjunto de datos

sintéticos en comparación con original

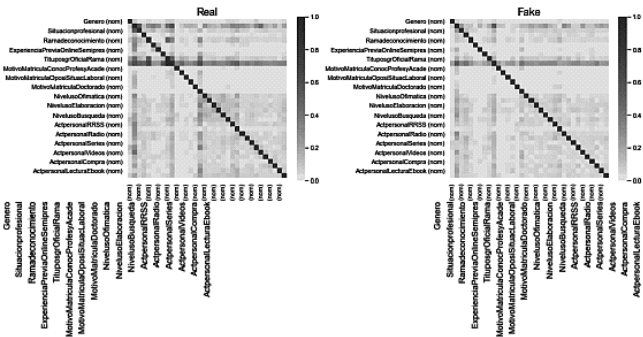
	151_1	151_2	151_3	151_4	151_5
KSTest		Inverted			0.9370
Kolmogorov-Smirnov statistic			D		
KSTestExtended		Inverted			0.9370
Kolmogorov-Smirnov statistic			D		
ContinuousKLDivergence					0.7813
Continuous Leibler Divergence			Kullback-		
Similarity Score					0.8850
0.9520	0.9477		0.9418		0.9466

0.9520	0.9477	0.9418	0.9466
0.8062	0.8109	0.7848	0.8043
0.9034	0.9021	0.8894	0.8991

Tras la comprobación de validez teórica y validez estadística en el grado de similitud de los conjuntos de datos originales y sintéticos, se optó por la realización de un análisis a partir de lo obtenido en los conjuntos de datos. Las primeras pruebas realizadas, aun incluyendo un grado de similitud alto, mostraban distorsiones cuando comparábamos los resultados obtenidos en el análisis de clúster bietápico en los resultados originales y en los conjuntos de datos sintéticos. La revisión y comprobación de las dificultades detectadas para la realización del análisis de clúster seleccionado, comprobó que, aun existiendo validez en la generación y fiabilidad en la validación entre conjuntos de datos, los conjuntos de datos sintéticos no mantenían la estructura de datos de los sujetos característica del conjunto original.

Figura 2.

Muestra de la correlación en nuestro conjunto de datos original y sintético



~ 147 ~

Anabel Bethencourt, Dagoberto Castellanos
RiiTE, Núm. 13 (2022), 138-156 Juan José Sosa y Manuel Area

Este hecho conllevó a nuevos procedimientos hacia las CopulaGAN-WAN en los que se incluyese en la librería en la que estábamos trabajando las correlaciones entre variables permitiendo el mantenimiento de los patrones e interrelaciones de las variables para cada sujeto. A modo de ejemplo de las correlaciones entre variables de nuestro conjunto de datos de las variables de “Nivel de uso

Herramientas digitales” por parte del alumnado se puede comprobar la siguiente tabla de datos.

Tabla 5.

Correlaciones subyacentes a la tabla de datos original entre las variables de Nivel de uso Herramientas digitales

Nivel uso Publicación	Nivel uso Ofimática Búsqueda	Nivel uso Organización Análisis de	Nivel uso Elaboración Datos
Nivel uso Ofimática		1	,463**
Nivel uso Organización	,463**		1
Nivel uso Elaboración	,183**	,359**	

Nivel uso	,158*	,374**
Publicación		
Nivel uso	,481**	,457**
Búsqueda		
Nivel uso	,166*	,213**
Análisis de Datos		

,183**	,158*	,481**	,166*
,359**	,374**	,457**	,213**
1	,471**	,201**	0,126
,471**	1	,248**	,186**
,201**	,248**	1	,294**
0,126	,186**	,294**	1

Como otro ejemplo de las correlaciones de nuestras variables, también se puede apreciar las correlaciones inherentes entre las variables de “Actividades personales de uso digital” en el alumnado.

Tabla 6.

Correlaciones subyacentes a la tabla de datos original entre las variables de Actividades personales de uso digital

	Frecuencia a las RRSS	Frecuencia a Internet	Frecuencia a escucha de la radio	Frecuencia a de televisión
Frecuencia las RRSS	1	,245**	,150*	0,033
Frecuencia Internet	,245**	1	,286**	,299**
Frecuencia escucha de la radio	,150*	,286**	1	,448**
Frecuencia de televisión	0,033	,299**	,448**	1
Frecuencia series	,232**	,360**	,233**	,378**

Frecuencia series películas online	Frecuencia videojuegos	Frecuencia videos en plataformas	Frecuencia de seguir conferencias	Frecuencia de compra online	Frecuencia videollamadas	Frecuencia lectura en ebook
,232**	,195**	,286**	,164*	,241**	,163*	,199**
,360**	0,112	,284**	,267**	0,141	,223**	,231**
,233**	-0,050	0,139	,228**	0,110	,152*	0,106
,378**	0,015	0,072	0,047	0,073	,173*	0,064
1	0,118	,335**	,220**	,244**	,192**	0,120

~ 148 ~

RiITE, Núm. 13 (2022), 138-156

películas online			
Frecuencia videojuegos	,195**	0,112	-0,050
Frecuencia videos en plataformas	,286**	,284**	0,139
Frecuencia de seguir	,164*	,267**	,228**

conferencias

Frecuencia de compra online ,241** 0,141 0,110

Frecuencia videollamadas ,163* ,223** ,152*

Frecuencia lectura en ebook ,199** ,231** 0,106

Implicaciones técnicas y prácticas de las Redes Adversarias Generativas a la Ciencia Abierta en Educación

0,015	0,118	1	,306**	,148*	,280**	0,075	,151*
0,072	,335**	,306**	1	,530**	,272**	,239**	,192**
0,047	,220**	,148*	,530**	1	,200**	,346**	,214**
0,073	,244**	,280**	,272**	,200**	1	,215**	,267**
,173*	,192**	0,075	,239**	,346**	,215**	1	,163*
0,064	0,120	,151*	,192**	,214**	,267**	,163*	1

Pese a que se muestre las variables correlacionadas dentro de una misma agrupación, es importante destacar que para que la estructura de tablas se mantenga durante el proceso de

creación de los datos sintéticos y mantengan una buena validez con respecto a los datos originales, se debe establecer las correlaciones de todas las variables presentes en el mismo conjunto de datos entre ellas, independientemente de que éstas formen parte de un mismo conjunto o bloque temático dentro del conjunto de datos. A partir de este proceso hacia las CopulaGAN-WAN se pudo obtener, de los análisis realizados, resultados más similares que incluyera la estructura de datos y, por lo tanto, que respete las tendencias y patrones que subyacen al conjunto de datos y que se manifiesta en las combinaciones de relaciones entre cada columna, coherentes con el conjunto de datos original.

4. DISCUSIÓN Y CONCLUSIONES

Las Redes Adversarias Generativas pueden favorecer la creación de datos sintéticos que permitan la protección de los datos y la difusión de las tablas entre la comunidad educativa y académica con la que poder realizar otros estudios o poner a

prueba y testear nuevos análisis. Esta característica es relevante para la Ciencia Abierta que incluye la puesta a disposición de librerías o conjuntos de datos con los que poder obtener las características fundamentales de los datos originales sin desatender las tareas en contra de la reidentificación de los agentes educativos involucrados ni compartir información protegida que incumpla un adecuado proceso de anonimización (Strcic et al., 2022; Abadal & Anglada, 2020; Alés, 2020).

Los resultados presentados muestran la equivalencia en grado de similitud de los datos originales con los datos creados a partir de los algoritmos, manteniendo una puntuación similar a los originales desde su proceso de generación de los datos, en su validación siguiente en las métricas estadísticas y, tras los ajustes de las variables correlacionadas. Realizar procedimientos de investigación en el que se incluyan los principios de la Ciencia Abierta también requiere la creación de estudios e investigaciones transparentes en el que se pueda ofrecer a la comunidad educativa los

conjuntos de datos sintéticos y, por lo tanto, anonimizados con fines académicos, investigadores o educativos. Este tipo de procedimientos, con tradición en las disciplinas técnicas y de ingeniería de datos, puede acercarse en la respuesta a la problemática de las pequeñas muestras en educación, así como a la protección de los datos que permita la transparencia y la

~ 149 ~

Anabel Bethencourt, Dagoberto Castellanos
RiiTE, Núm. 13 (2022), 138-156 Juan José Sosa y Manuel
Area

ética del proceso investigador, relevante en el campo de la Tecnología Educativa y en la Ciencia Abierta. Desarrollar estrategias más éticas y transparentes en el ámbito de la Ciencia Abierta y de la Tecnología Educativa implica el trabajo conjunto y colaborativo en distintos campos disciplinares que permita dar respuesta a las complejidades inherentes al campo educativo y del tratamiento de datos e información (Parti &

Szigeti, 2021; Bethencourt-Aguilar et al., 2021). Además, dentro de la investigación en Tecnología Educativa, las GAN pueden aportar la técnica que permita la reutilización, el intercambio y la difusión de información, recursos digitales y conjuntos de datos sobre nuestra línea que nos permita la toma de decisión hacia una transformación digital alineada con los principios de la Ciencia Abierta, es decir, más accesible, colaborativa y en beneficio de la sociedad (Alhadad et al., 2018; Bethencourt-Aguilar et al., 2021ayb; Alés, 2020). En el ámbito de la investigación educativa, este nuevo procedimiento está estrechamente relacionado con la Tecnología Educativa porque en ésta se 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.

Entre las principales problemáticas encontradas en el proceso de creación de las tablas sintéticas, se destaca la importancia en la creación de las tablas donde se

mantenga la relación entre las variables. Este hecho es fundamental para la investigación en educación, donde a menudo en los análisis resulta relevante advertir los patrones, percepciones y perfiles emergentes en los datos. Si no se realiza una cuidada creación de datos sintéticos manteniendo las estructuras de las relaciones (o correlaciones) entre variables, se puede perder información relevante de las características de los sujetos y de los registros subyacentes en el cruce de datos. De este modo, el mantenimiento de la estructura de las tablas permite generar conjuntos de datos sintéticos que mantengan una similitud también en sus características más subjetivas y pueda realizarse análisis de datos que incluyan la combinación de variables. Por ello, es relevante mostrar las posibilidades de replicar conjuntos de datos entrenados para la creación de otros sintéticos que incluya la misma estructura de datos original manifestada en las diversas combinaciones de relaciones entre cada columna de nuestro conjunto de datos del que parte nuestro estudio. El mantenimiento de estas estructuras es, en sí

misma, lo que permite ofrecer resultados de investigación que sea replicable con los datos originales y que no incluya una correlación nula y se muestren independientes en los conjuntos.

Las líneas futuras de investigación conducen a analizar hasta qué punto se puede suplir o aumentar el conjunto de datos, contrastando los resultados en el conjunto original y la muestra sintética ante los mismos procedimientos de análisis. Independientemente de que los datos den resultados positivos, negativos o regulares, la adecuación en sustitución de una muestra original con una muestra sintética creada de modo artificial a través de estos algoritmos indicaría que estos resultados son los mismos en los sintéticos como en los originales. Analizar los resultados obtenidos de un conjunto de datos sintéticos extendidos podría ser relevante ante la ausencia o dificultades de muestra de nuestra investigación educativa y, además, para verificar si emergen nuevos resultados no contemplados por la escasez de muestra (Creswell et al., 2018). En otras palabras, el uso de las GAN ofrecen ventajas para

la investigación educativa, al posibilitar la creación de datos sintéticos que incrementan el número de registros muestrales disponibles. Las ventajas de todo ello son evidentes. Además de incrementar los tamaños muestrales con un coste mínimo, lo que facilita la aplicación de técnicas de análisis alternativas, no aplicables cuando las

~ 150 ~

RiiTE, Núm. 13 (2022), 138-156

Implicaciones técnicas y prácticas de las Redes Adversarias

Generativas a la Ciencia Abierta en Educación

muestras son pequeñas y la anonimización de datos, disponer y compartir estas bases de datos facilitan la ciencia abierta y la aplicabilidad de estrategias de análisis de datos basados en aprendizaje profundo e inteligencia artificial, aún en desarrollo en el campo de la educación y de la tecnología educativa. Independientemente de estas ventajas, es importante considerar que las limitaciones de esta técnica están sujetas a las limitaciones de calidad de la muestra original, además de a un uso

éticamente responsable (como con cualquier otro procedimiento de muestreo o de análisis numérico). Lo que en este trabajo se propone debe interpretarse, simplemente, como una alternativa o posibilidad más a considerar, en el ámbito de la investigación en educación, siendo conscientes de sus ventajas y de sus limitaciones, entre las que la sofisticación y complejidad técnica del procedimiento no es una cuestión menor. En este sentido, entre las limitaciones técnicas podemos destacar la selección del algoritmo más adecuado, el tiempo de entrenamiento y validación, y las complejidades en la explicación de sus resultados, que serían elementos que pueden constituir limitaciones para los profesionales de la educación.

La investigación en general y con ella la educativa, parte o se basa también en los datos, en la cantidad y calidad de ellos, en el poder de su replicación, sus relaciones y en principios como la privacidad de los mismos. Las GAN posibilitan estas acciones asegurando la privacidad de los datos e impulsando

el desarrollo de técnicas y, por lo tanto, de análisis emergentes relevantes para los emergentes métodos de investigación en Tecnología Educativa. La Tecnología Educativa debe impulsar prácticas más abiertas y colaborativas en las distintas fases y etapas del proceso de investigación y para ello, la inserción de las Redes Adversarias Generativas podría ser relevante para favorecer métodos investigadores más éticos y transparentes, de acuerdo con los principios de la Ciencia Abierta.

5. ENLACES

En este apartado compartimos el cuestionario donde aparecen las variables descritas en el artículo, así como el conjunto de datos sintéticos que al término de este estudio obtuvimos. Acceso al cuestionario: [enlace](#)

Acceso al conjunto sintético: [enlace](#)

6. RECONOCIMIENTOS O FINANCIACIÓN

Esta investigación está financiada por el Ministerio de Ciencia, Innovación y Universidades a través de un contrato en el Programa de Formación de Profesorado Universitario (FPU) en el Departamento de Didáctica e Investigación Educativa de la Facultad de Educación de la Universidad de La Laguna (FPU19/04821). La doctoranda es Anabel Bethencourt Aguilar, el director Manuel Area Moreira y los codirectores Juan José Sosa Alonso y Dagoberto Castellanos Nieves. La tesis doctoral lleva por nombre “La transformación digital de la enseñanza universitaria. Análisis y propuestas para las titulaciones de posgrado en la Universidad de La Laguna”.

Esta investigación surge en el grupo de investigación e innovación EDULLAB: Laboratorio de Educación y Nuevas Tecnologías de La Universidad de La Laguna. El equipo de investigación

agradece a la Escuela de Doctorado y Estudios de Postgrado, al Gabinete de Planificación y Comunicación, al Vicerrectorado de Agenda Digital, Modernización y Campus Central, al Vicerrectorado de Innovación Docente, Calidad y Campus Anchieta y la Unidad de Docencia Virtual (UDV) de la Universidad de La Laguna por su apoyo y colaboración en la investigación que enmarca la tesis doctoral.

7. REFERENCIAS BIBLIOGRÁFICAS

Abadal, E., & Anglada, L. (2020). Ciencia abierta: cómo han evolucionado la denominación y el concepto. *Anales de Documentación*, 23(1). <https://doi.org/10.6018/analesdoc.378171>

Alés, N. S. (2020). La Ciencia y Educación Abierta como movimientos articuladores de la investigación, la tecnología y la innovación:

Experiencias del proyecto de Acceso Abierto de la Facultad de Comunicación de la Universidad de La Habana. *Revista Publicando*, 7(27), 65-72.

Alhadad, S. S. J., Searston, R. A., & Lodge, J. M. (2018). Interdisciplinary open science: What are the implications for educational technology research?. In M. Campbell, J. Willems, C. Adachi, D. Blake, I. Doherty, S. Krishnan, S. Macfarlane, L. Ngo, M. O'Donnell, S. Palmer, L. Riddell, I. Story, H. Suri & J. Tai (Eds.), *Open Oceans: Learning without borders*. Proceedings ASCILITE 2018 Geelong, (pp. 303-308).

Al-Qizwini, M., Barjasteh, I., Al-Qassab, H., & Radha, H. (2017). Deep learning algorithm for autonomous driving using GoogLeNet. *2017 IEEE Intelligent Vehicles Symposium (IV)*, 89-96.
<https://doi.org/10.1109/IVS.2017.7995703>

Barua, S., Islam, M. M., Yao, X., & Murase, K. (2014). MWMOTE--Majority Weighted Minority Oversampling Technique for Imbalanced Data Set Learning. *IEEE Transactions on Knowledge and Data Engineering*, 26(2),

Bethencourt Aguilar, A., Castellanos Nieves, D., Sosa Alonso, J. J., & Area Moreira, M. (2022).

Synthetic student dataset on levels of use of digital tools and frequency of personal activities with ICTs, Mendeley Data, v1, <http://dx.doi.org/10.17632/rwz59sxtpn.1>

Bethencourt-Aguilar, A., Area-Moreira, M., Sosa-Alonso, J. J., & Castellano-Nieves, D. (2021). The digital transformation of postgraduate degrees. A study on academic analytics at the University of La Laguna. *2021 XI International Conference on Virtual Campus (JICV)*, 1-4.

<https://doi.org/10.1109/JICV53222.2021.9600311>

Bethencourt-Aguilar, A., Sosa-Alonso, J. J., Castellanos-Nieves, D. C., & Area-Moreira, M. (2021). Uso del campus virtual y el rendimiento académico del alumnado: Análisis antes, durante y después del impacto de la Covid-19 en la educación superior. *InnoEduca Tic 2021: Libro de Actas de las VIII Jornadas Iberoamericanas de Innovación Educativa*

en el ámbito de las TIC y las TAC Las Palmas de Gran Canaria, 18 y 19 de noviembre de 2021, 2021, ISBN 978-84-0935708-6, págs. 293-297. <https://dialnet.unirioja.es/servlet/articulo?codigo=8227886>

Bishop, C. M. (1995). Training with Noise is Equivalent to Tikhonov Regularization. *Neural Computation*, 7(1), 108-116. <https://doi.org/10.1162/neco.1995.7.1.108>

Burgos, D. (2020). *Radical Solutions and Open Science: An Open Approach to Boost Higher Education*. Springer

Nature. <https://doi.org/10.1007/978-981-15-4276-3>

~ 152 ~

RiiTE, Núm. 13 (2022), 138-156

Implicaciones técnicas y prácticas de las Redes Adversarias

Generativas a la Ciencia Abierta en Educación

Chawla, N. V., Bowyer, K. W., Hall, L. O., & Kegelmeyer, W. P. (2002). SMOTE: Synthetic Minority

Over-sampling Technique. *Journal of Artificial Intelligence Research*, 16, 321-357.
<https://doi.org/10.1613/jair.953>

Conole, G., & Brown, M. (2018). Reflecting on the Impact of the Open Education Movement. *Journal of Learning for Development*, 5(3).
<https://doi.org/10.56059/jl4d.v5i3.314>

Creswell, A., White, T., Dumoulin, V., Arulkumaran, K., Sengupta, B., & Bharath, A. A. (2018). Generative Adversarial Networks: An Overview. *IEEE Signal Processing Magazine*, 35(1), 53-65. <https://doi.org/10.1109/MSP.2017.2765202>

Douzas, G., & Bacao, F. (2017). Self-Organizing Map Oversampling (SOMO) for imbalanced data set learning. *Expert Systems with Applications: An International Journal*, 82(C), 40-52.
<https://doi.org/10.1016/j.eswa.2017.03.073>

Douzas, G., & Bacao, F. (2018). Effective data generation for imbalanced learning using conditional generative adversarial networks.

Expert Systems with Applications, 91, 464-471.
<https://doi.org/10.1016/j.eswa.2017.09.030>

DeRouin, E. & Brown, J. (1991). Neural Network Training on Unequally Represented Classes. *Intelligent Engineering Systems through Artificial Neural Networks*, 135-140.

Fiore, U., De Santis, A., Perla, F., Zanetti, P., & Palmieri, F. (2019). Using generative adversarial networks for improving classification effectiveness in credit card fraud detection. *Information Sciences*, 479, 448-455. <https://doi.org/10.1016/j.ins.2017.12.030>

Fressoli, J. M., & Arza, V. (2018). Los desafíos que enfrentan las prácticas de ciencia abierta. *Teknokultura. Revista de Cultura Digital y Movimientos Sociales*, 15(2). <https://doi.org/10.5209/TEKN.60616>

González-Pérez, L. I., Ramírez-Montoya, M. S., & García-Peñalvo, F. J. (2022). Technological Enablers 4.0 to Drive Open Science and Education:

Input to UNESCO Recommendations. *RIED-Revista Iberoamericana de Educación a Distancia*, 25(2), 23-48.
<https://doi.org/10.5944/ried.25.2.33088>

Goodfellow, I., Pouget-Abadie, J., Mirza, M., Xu, B., Warde-Farley, D., Ozair, S., Courville, A., & Bengio, Y. (2014). Generative Adversarial Nets. *Advances in Neural Information Processing Systems*, 2
7. <https://papers.nips.cc/paper/2014/hash/5ca3e9b122f61f8f06494c97b1afccf3-Abstract.html>

Gou, C., Wu, Y., Wang, K., Wang, F.-Y., & Ji, Q. (2016). Learning-by-synthesis for accurate eye detection. *2016 23rd International Conference on Pattern Recognition (ICPR)*, 3362-3367.
<https://doi.org/10.1109/ICPR.2016.7900153>

Han, H., Wang, W.-Y., & Mao, B.-H. (2005). Borderline-SMOTE: A New Over-Sampling Method in Imbalanced Data Sets Learning. En D.-S. Huang, X.-P. Zhang, & G.-B. Huang (Eds.), *Advances in Intelligent Computing* (pp. 878-887).

He, H., Bai, Y., Garcia, E. A., & Li, S. (2008). ADASYN: Adaptive synthetic sampling approach for imbalanced learning. *2008 IEEE International Joint Conference on Neural Networks (IEEE World Congress on Computational Intelligence)*, 1322-1328
<https://doi.org/10.1109/ijcnn.2008.4633969>

~ 153 ~

Anabel Bethencourt, Dagoberto Castellanos
RiiTE, Núm. 13 (2022), 138-156 Juan José Sosa y Manuel Area

Hinton, G. E., & Salakhutdinov, R. R. (2006). Reducing the dimensionality of data with neural networks. *Science (New York, N.Y.)*, 313(5786), 504-507.
<https://doi.org/10.1126/science.1127647>

Jones, N. (2015). Artificial-intelligence institute launches free science search engine. *Nature*.
<https://doi.org/10.1038/nature.2015.18703>

Li, D.-C., & Fang, Y.-H. (2009). A non-linearly virtual sample generation technique using group discovery and parametric equations of hypersphere. *Expert Systems with Applications*, 36(1), 844-85

1. <https://doi.org/10.1016/j.eswa.2007.10.029>

Li, D.-C., & Lin, Y.-S. (2006). Using virtual sample generation to build up management knowledge in the early manufacturing stages. *European Journal of Operational Research*, 175(1), 413-434. <https://doi.org/10.1016/j.ejor.2005.05.005>

Li, J., Monroe, W., Shi, T., Jean, S., Ritter, A., & Jurafsky, D. (2017). *Adversarial Learning for Neural Dialogue Generation* <https://doi.org/10.48550/arXiv.1701.06547>

Logan, J. A. R., Hart, S. A., & Schatschneider, C. (2021). Data Sharing in Education Science. *AERA Open*, 7, Cornell University. <https://doi.org/10.1177/23328584211006475>

Mohamed, A., Dahl, G. E., & Hinton, G. (2012). Acoustic Modeling Using Deep Belief Networks. *IEEE Transactions on Audio, Speech, and Language Processing*, 20(1), 14-22.
<https://doi.org/10.1109/TASL.2011.2109382>

Parti, K., & Szigeti, A. (2021). The Future of Interdisciplinary Research in the Digital Era: Obstacles and Perspectives of Collaboration in Social and Data Sciences - An Empirical Study. *Cogent Social Sciences*, 7(1). <https://doi.org/10.1080/23311886.2021.1970880>

Pascual, S., Bonafonte, A., & Serrà, J. (2017). *SEGAN: Speech Enhancement Generative Adversarial Network*, Cornell University.
<https://doi.org/10.48550/arXiv.1703.09452>

Peset, F., & Millán González, L. (2017). *Ciencia abierta y gestión de datos de investigación: RDM*. Ediciones Trea.

Radford, A., Metz, L., & Chintala, S. (2015).

Unsupervised Representation Learning with Deep Convolutional Generative Adversarial Networks, Computer Science, abs/1511.06434.

Ramírez-Montoya, M. S., McGreal, R., & Agbu, J.-F. O. (2022). Complex Digital Horizons in the Future of Education 4.0: Insights from UNESCO Recommendations. *RIED-Revista Iberoamericana de Educacion a Distancia*, 25(2), 09-21.
<https://doi.org/10.5944/ried.25.2.33843>

UNESCO (2021) Recomendación de la UNESCO sobre la Ciencia Abierta—*UNESCO Biblioteca Digital*. Recuperado 2 de noviembre de 2022, de https://unesdoc.unesco.org/ark:/48223/pf0000379949_spa

Santana, E., & Hotz, G. (2016a). Learning a Driving Simulator. En *ArXiv e-prints*.
<https://ui.adsabs.harvard.edu/abs/2016arXiv160801230S>

Strcic, J., Civljak, A., Glozinic, T., Pacheco, R. L., Brkovic, T., & Puljak, L. (2022). Open data and data sharing in articles about COVID-19 published

in preprint servers medRxiv and bioRxiv.

Scientometrics, 127(5),

2791-2802. <https://doi.org/10.1007/s11192-022-04346-1>

~ 154 ~

RiiTE, Núm. 13 (2022), 138-156

Implicaciones técnicas y prácticas de las Redes

Adversarias

Generativas a la Ciencia Abierta en Educación

Theodoridis, S., & Koutroumbas, K. (2006). *Pattern Recognition, Third Edition*. Academic Press, Inc.

Van der Zee, T., & Reich, J. (2018). Open Education Science. *AERA Open*, 4(3).
<https://doi.org/10.1177/2332858418787466>

Van Dijk, W., Schatschneider, C., & Hart, S. A. (2021). Open Science in Education Sciences. *Journal of Learning Disabilities*, 54(2),

139-152. <https://doi.org/10.1177/0022219420945267>

Wang, K., Gou, C., Duan, Y., Lin, Y., Zheng, X., &

Wang, F.-Y. (2017). Generative adversarial networks: Introduction and outlook. *IEEE/CAA Journal of Automatica Sinica*, 4(4), 588-598. <https://doi.org/10.1109/JAS.2017.7510583>

Wang, L., & Sng, D. (2015). *Deep Learning Algorithms with Applications to Video Analytics for A Smart City: A Survey*, Cornell University. <https://doi.org/10.48550/arXiv.1512.03131>

Xie, Z., Jiang, L., Ye, T., & Li, X. (2015). A Synthetic Minority Oversampling Method Based on Local Densities in Low-Dimensional Space for Imbalanced Learning. En M. Renz, C. Shahabi, X. Zhou, & M. A. Cheema (Eds.), *Database Systems for Advanced Applications* (pp. 3-18). Springer International Publishing. https://doi.org/10.1007/978-3-319-18123-3_1

Zhou, Z.-H., & Jiang, Y. (2004). NeC4.5: Neural ensemble based C4.5. *IEEE Transactions on Knowledge and Data Engineering*, 16(6), 770-773. <https://doi.org/10.1109/TKDE.2004.11>

Zhuang, Y., Wu, F., Chen, C., & Pan, Y. (2017). Challenges and opportunities: From big data to knowledge in AI 2.0. *Frontiers of Information Technology & Electronic Engineering*, 18(1), 3-14. <https://doi.org/10.1631/FITEE.1601883>

INFORMACIÓN SOBRE LOS AUTORES

Anabel Bethencourt-Aguilar

Universidad de La Laguna

Contratada Predoctoral en el Programa de Formación del Profesorado Universitario del Ministerio de Universidades del Gobierno de España, en el Departamento de Didáctica e Investigación Educativa de la Facultad de Educación, Universidad de La Laguna. Miembro del grupo de investigación e innovación EDULLAB (Laboratorio de Nuevas Tecnologías y Educación) desde 2017 hasta la actualidad y miembro de la Red Universitaria de Investigación e Innovación Educativa (REUNI+D). Miembro de la Cátedra Fundación Mapfre Guanarteme de Tecnología y Educación de la Universidad de La Laguna (TECNOEDU).

Dagoberto Castellano-Nieves

Universidad de La Laguna

Profesor Titular de Universidad en el departamento de Ingeniería Informática y de Sistemas de la ULL. Su ámbito de investigación se ha centrado a diversos temas relacionados con el aprendizaje profundo, análisis predictivos, computación de altas prestaciones, representación del conocimiento, sistemas recomendadores, localización inteligente y, con las nuevas tendencias en educación y las tecnologías educativas. Forma parte de los grupos de investigación "Grupo de Computación Inteligente" (GCI) y "Grupo de Computación de Altas Prestaciones (GCAP), ambos de la Universidad de La Laguna.

~ 155 ~

Anabel Bethencourt, Dagoberto Castellanos

RiiTE, Núm. 13 (2022), 138-156 Juan José Sosa y Manuel Area

Juan José Sosa-Alonso

Universidad de La Laguna

Profesor Ayudante Doctor en el departamento de Didáctica

e Investigación Educativa de la ULL. Su ámbito de investigación se ha centrado a diversos temas relacionados con evaluación psicopedagógica, atención al alumnado con NEAE, evaluación institucional, asesoramiento a familias y, más recientemente, con las nuevas tendencias en educación y la tecnología educativa. Forma parte de los grupos de investigación consolidados "Laboratorio de Educación y Nuevas Tecnologías" (EDULLAB) y "Poder, Saber y Subjetividad en la Educación" (POSASUE), ambos de la Universidad de La Laguna.

Manuel Area-Moreira

Universidad de La Laguna

Dr. en Pedagogía y Catedrático de la Universidad de La Laguna ULL (Islas Canarias. España) en el Departamento de Didáctica e Investigación Educativa en la Facultad de Educación. Su ámbito de investigación y docencia es la Tecnología Educativa (Cultura digital y educación, Enseñanza con medios y tecnologías, eLearning, Alfabetización y TIC, Políticas educativas y ciudadanía digital, ...). Es Investigador Principal del Laboratorio de Educación y Nuevas Tecnologías (EDULLAB) y Director de la Cátedra de Tecnología y Educación de Mapfre-Guanarteme en la ULL.



Los textos publicados en esta revista están sujetos a una licencia de Reconocimiento 4.0 España de Creative Commons. Puede copiarlos, distribuirlos, comunicarlos públicamente y hacer obras derivadas siempre que reconozca los créditos de las obras (autoría, nombre de la revista, institución editora) de la manera especificada por los autores o por la revista. La licencia completa se puede consultar en: [Licencia](#)

[Creative Commons Atribución-NoComercial-Compartir por igual 4.0 Internacional.](#)