



Aplicaciones de Redes Neuronales Transformers

Mónica Ramírez ¹ ; Marcelo Gudiño ² 
0000-0003-3103-6797 ¹; 0009-0002-0297-9470 ²
monica.ramirez@itq.edu.ec¹; migsblues1@hotmail.com²
Corresponsal: monica.ramirez@itq.edu.ec; Telf.: +593 9842430091

Resumen: Las redes neuronales transformers han transformado áreas clave como el procesamiento del lenguaje natural (PLN), la visión por computadora y la predicción en diversos campos, incluyendo educación, finanzas y análisis de imágenes. Su capacidad para procesar grandes volúmenes de datos y captar patrones complejos ha impulsado su uso en aplicaciones como traducción automática, generación de texto, reconocimiento facial y predicción de series temporales. Este artículo introduce el funcionamiento de los Transformers, destacando sus fundamentos teóricos y diferencias con otras arquitecturas como las redes neuronales recurrentes (RNN) y convolucionales (CNN). Además, se analizan sus aplicaciones en distintos sectores, resaltando avances en precisión, velocidad de procesamiento y eficiencia computacional. A pesar de sus ventajas, estos modelos enfrentan desafíos como el alto costo computacional, la necesidad de grandes volúmenes de datos y la dificultad en la interpretabilidad de sus predicciones. El estudio también explora tendencias futuras, incluyendo mejoras en accesibilidad, eficiencia y adaptación a nuevos contextos. Con innovaciones en marcha, los Transformers continúan consolidándose como una tecnología clave en la evolución de la inteligencia artificial y el aprendizaje profundo.

Palabras clave: Inteligencia Artificial; transformers, redes neuronales, deep learning, predicción.

Abstract: Transformer neural networks have transformed key areas such as natural language processing (NLP), computer vision, and prediction in diverse fields, including education, finance, and image analysis. Their ability to process large volumes of data and capture complex patterns has driven their use in applications such as machine translation, text generation, and facial recognition. This article introduces how Transformers work, highlighting their theoretical foundations and differences with other architectures such as recurrent neural networks (RNNs) and convolutional neural networks (CNNs). It also analyzes their applications in various sectors, highlighting advances in accuracy, processing speed, computational efficiency and time series prediction. Despite their advantages, these models face challenges such as high computational costs, the need for large volumes of data, and



Cita: Ramírez. & Gudiño. Aplicaciones de redes neuronales Transformers. Revista DOXA ITQ, 3(1), 002.

Recibido: 07/08/2024

Aceptado: 07/11/2024

Publicado: 01/02/2025

Keyerman Toapanta C. M.Sc. Editor en jefe, Revista DOXA ITQ Quito, Ecuador.

Nota del editor: DOXA Editorial mantiene neutralidad respecto a cualquier reclamo legal derivado del contenido publicado en la Revista DOXA ITQ. La responsabilidad por la información recae completamente en los autores.

the difficulty in interpreting their predictions. The study also explores future trends, including improvements in accessibility, efficiency, and adaptation to new contexts. With ongoing innovations, Transformers continue to establish themselves as a key technology in the evolution of artificial intelligence and deep learning.

Keywords: Artificial intelligence; transformers; neural networks; deep learning; prediction



1. Introducción

En los últimos 10 años, la inteligencia artificial (IA) ha avanzado notablemente gracias al desarrollo de modelos de aprendizaje profundo. En particular, aquellas redes neuronales llamadas Transformers han surgido como una potencia arquitectónica que ha transformado diversas áreas del conocimiento; en particular, el área del procesamiento del lenguaje natural (PLN) y la visión por computadora. Su eficiencia en la manipulación de secuencias de datos, sin estar sujeto a las restricciones que imponen las redes neuronales recurrentes (RNN) ha permitido avances tecnológicos sin precedentes, aplicadas a tareas tales como la traducción automática, la generación de texto y el análisis de imágenes. Esta tecnología emergente ha propiciado el interés creciente por la aplicación de Transformers en diferentes campos científicos y comerciales.

El término Transformers se refiere a un modelo fundamentado en sistemas de atención que es capaz de procesar información en paralelo, así como de establecer relaciones entre los elementos de una secuencia sin tener que forzar un procesamiento estrictamente de tipo secuencial. En contraste con las RNN y sus variantes, como las LSTM (Long Short-Term Memory) y GRU (Gated Recurrent Unit), los Transformers han mostrado un mejor rendimiento en la captura de dependencias a largo plazo, así como también en la optimización de recursos computacionales, por lo que han podido ser implementados en modelos de última generación como BERT (Bidirectional Encoder Representations from Transformers) y GPT (Generative Pre-trained Transformer), los cuales han conseguido redefinir el

aprendizaje automático y la generación de lenguaje natural.

El alcance de los Transformers no se restringe solamente al procesamiento del lenguaje natural. Su flexibilidad ha permitido, la utilidad en otras áreas del aprendizaje profundo, como la predicción de series temporales, el análisis de datos biomédicos y la detección de anomalías en señales de audio. En la visión por computadora, los Vision Transformers (ViTs) han superado a las redes neuronales convolucionales (CNNs) en muchas tareas, pues su capacidad de modelar relaciones espaciales ha dado resultados mucho más eficientes. Estas aplicaciones han hecho de los Transformers un pilar de la inteligencia artificial moderna, con implicaciones en tareas de automatización, toma de decisiones y optimización de procesos en diferentes sectores.

A pesar de su éxito, la implementación de Transformers presenta importantes retos. Por un lado, poseen una enorme necesidad de datos y, por otro lado, requiere una enorme capacidad computacional, lo que establece una serie de barreras de accesibilidad y sostenibilidad. Por otra parte, la interpretación de estos modelos continúa siendo uno de los puntos de estudio hoy en día, ya que la complejidad de los modelos hace que sea difícil comprender los procesos de toma de decisiones dentro de la red. Por eso, existen varias propuestas de técnicas que procuran incrementar la proporción de eficiencia de los Transformers, ya sea mediante la reducción de los parámetros, la mejora de la arquitectura o el aprendizaje auto-supervisado.

Las redes neuronales Transformers están teniendo un impacto y una evolución muy rápida. Por lo tanto, el objetivo de este

artículo es desarrollar una revisión en profundidad del uso de las redes neuronales Transformers en los diferentes ámbitos del aprendizaje profundo, teniendo en cuenta los fundamentos teóricos en que se basan, los principales modelos de los que se compone y sus aplicaciones de mayor tamaño, así como las problemáticas actuales de su aplicación. Del mismo modo, a partir de este estudio se pretende ofrecer una visión global del estado del arte en este ámbito para que dé lugar a propuestas de futuras investigaciones y desarrollos en el ámbito de la tecnología.

2. Materiales y métodos

Este trabajo de revisión se apoya en la recopilación de diferentes orígenes de literatura científica y de documentos técnicos de interés sobre el empleo de las denominadas redes neuronales Transformers para el aprendizaje automático, haciendo hincapié en el procesamiento de lenguaje natural y en la visión por computadora. Se procedió a la recopilación de artículos científicos de bases de datos concretas como Google académico, IEEE Xplore, SpringerLink y Scopus, dándole preferencia a publicaciones recientes y de interés específico.

La búsqueda de información se llevó a cabo utilizando palabras clave y aplicando criterios de inclusión y exclusión para seleccionar sólo estudios con validación experimental. Los datos extraídos incluyeron modelos utilizados, conjunto de datos empleados y métricas de rendimiento.

3. Redes Neuronales y Aprendizaje Profundo

Existen algunas limitaciones con las RNN, ya que tienen dificultad para captar relaciones o dependencias en secuencias muy largas, así que se empezó a desarrollar mecanismos

como las LSTM (Long Short-Term Memory) y las GRU (Gated Recurrent Unit) las arquitecturas de estas redes neuronales incluyen puertas que permiten regular el paso de la información, para así manejar de mejor forma los contextos más complejos y de mayor extensión (Hochreiter & Schmidhuber, 1997).

Las arquitecturas planteadas han tenido un papel esencial en tareas desde el reconocimiento del habla hasta la traducción automática (Cho et al., 2014).

Puede decirse que las redes neuronales son modelos inspirados en el cerebro humano que han sido diseñados para la resolución de problemas complejos a partir del descubrimiento de patrones en los datos. Las arquitecturas fundamentales incluyen el Perceptrón Multicapa (MLP), las Redes Neuronales Convolucionales (CNN) y las Redes Neuronales Recurrentes (RNN). Las versiones más modernas, como las denominadas Transformers, han ampliado el campo de aplicación de las redes neuronales al facilitar el procesamiento eficiente de la información de datos secuenciales y dependencias a largo plazo (Vaswani et al., 2017).

3.1. Redes Transformers

Si bien el término Transformers está relacionado con los robots este término no tiene ninguna relación con ellos, más bien proviene del hecho que el modelo transforma y procesa información de manera eficiente.

Las redes Transformers fueron presentadas por Vaswani et al. (2017) como un avance notable en el campo del procesamiento del lenguaje. La diferencia más sustancial con las RNN es que los transformers no procesan la información de manera secuencial y, por tanto, no están sometidos a la strictness de la

secuencialidad. De hecho, los transformers son procesadores paralelos; su secreto está en los mecanismos de atención, que aprovechan la posibilidad de trabajar con toda la entrada de una sola vez, en lugar de usarla de forma encadenada durante el tiempo. Por lo tanto, se puede afirmar que las arquitecturas de aprendizaje profundo denominadas "Transformers" son un tipo de red neuronal que utiliza mecanismos de atención para trabajar con secuencias de datos. Estos datos pueden ser las palabras en una oración, que luego son utilizadas por la red para crear un contexto a partir del cual pueden hacer inferencias de tipo estructural.

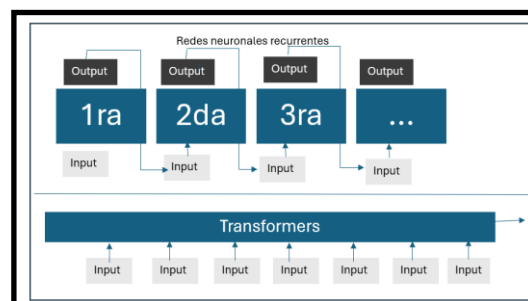
Los modelos Transformers, a diferencia de las redes neuronales recurrentes (RNN) tradicionales y sus variantes como las LSTM, pueden procesar secuencias de entrada de forma paralela. También son mejores para manejar las dependencias a largo plazo. Esto es así porque utilizan capas de atención multicabeza y, por lo tanto, pueden captar relaciones entre elementos que están distantes en la secuencia.

Se sabe que los Transformers han sido muy eficaces en diversas tareas de aprendizaje automático, particularmente en el procesamiento de lenguaje natural. También han tenido notable éxito en otras áreas, como las siguientes: visión computacional, análisis en el procesamiento de señales de audio, entre otros. Los modelos Transformers, a diferencia de las redes neuronales recurrentes (RNN) tradicionales, pueden procesar secuencias de entrada de forma paralela. Esto es así porque utilizan capas de atención multicabeza y, por lo tanto, pueden captar relaciones entre elementos que están distantes en la secuencia. Por ejemplo, en su

artículo, Won et al. (2023) indica que el uso de Transformers para modelar datos demográficos como (género, nacionalidad, año de graduación de secundaria) y académicos como (participación en actividades, evaluaciones de curso) ha permitido predecir la deserción con una precisión notablemente frente a técnicas anteriores. De esta información se fundamenta que los modelos de machine learning para predecir la deserción estudiantil son útiles para abordar este problema.

Estas redes neuronales tienen la capacidad de aprender el contexto y el significado al analizar las relaciones en datos secuenciales, como las palabras en una oración por ejemplo o la secuencia de procesos. Inicialmente, se utilizaron para la traducción de un idioma a otro. Pero en la actualidad gracias a los Transformers y su mecanismo de atención, surgieron los grandes modelos generadores de texto, como el GPT4.0 y BERT. Estos modelos ahora pueden ser entrenados de manera más eficiente gracias al aprovechamiento del paralelismo que se logra mediante el uso de GPUs.

Figura 1
Comparación entre redes neuronales recurrentes y redes transformers.



Nota. Funcionamiento de redes neuronales recurrentes y redes Transformers. *Fuente.* Elaboración propia.

Las principales características de la arquitectura Transformer que se pueden citar según Vaswani, et al. (2017) son:

- La autoatención es un mecanismo fundamental de los Transformers que permite al modelo enfocarse en diferentes partes de la secuencia de entrada para cada posición de salida. Este mecanismo elabora una representación de cada término, en función de todos los demás términos de la secuencia. Es un método que permite captar mejor las dependencias a largo plazo entre las palabras de una misma secuencia.
- Paralelización: A diferencia de las RNN, que procesan las secuencias de datos de forma secuencial, el Transformer puede procesar todas las posiciones de la secuencia de entrada en paralelo. Esto es posible gracias a la eliminación de las dependencias temporales y, desde luego, a la naturaleza del Transformer como tal.
- Codificadores y decodificadores. La arquitectura Transformer está constituida por un apilamiento de codificadores y otro de decodificadores. Cada codificador y cada decodificador tiene capas que contienen los mecanismos de autoatención y capas feedforward (que se puede traducir como retroalimentación).

Atención Multi-Cabeza: Esta parte del modelo le permite concentrarse en diferentes representaciones de las posiciones de la secuencia. Está mejorando la capacidad del modelo para captar varias características de los datos.

3.1.1. Arquitectura de Redes Transformers

La arquitectura básica de un Transformer incluye:

- Embeddings: Son representaciones vectoriales de palabras. Mecanismo de Atención: Calculan la importancia relativa de cada palabra en el contexto de una oración.
- Capas Feedforward: Aplican transformaciones no lineales para mejorar la capacidad de aprendizaje.
- Codificadores y Decodificadores: Son componentes principales que procesan y generan datos respectivamente (Vaswani et al., 2017).

Figura 2
La Arquitectura del Transformer.



Nota. Componentes principales de modelos transformers: Encoders y Decoders. Imagen tomada de Aprende Machine Learning (2022).

Fuente:

<https://www.aprendemachinlearning.com/como-funcionan-los-transformers-espanol-nlp-gpt-bert/>

3.2. Modelos preentrenados

En el campo del procesamiento de lenguaje natural sobresale el modelo BERT que utiliza una arquitectura basada en codificadores, este modelo es preentrenado en tareas como el modelado de lenguaje y la

predicción de palabras enmascaradas (Devlin et al, 2018).

3.2.1. BERT (Bidirectional Encoder Representations from Transformers)

Es un modelo que se preentrena y que destaca en la tarea de entender el lenguaje natural. BERT se entrena de manera bidireccional. Esto significa que el modelo tiene en cuenta tanto el contexto anterior como el posterior de una palabra en una oración. Esta es la razón por la que BERT es un modelo que profundiza en la comprensión de las relaciones entre palabras mejorando el rendimiento en tareas de NLP. Este modelo está compuesto únicamente por codificadores basados en la arquitectura de transformers. Cada codificador incluye múltiples capas de atención y redes conectadas.

El modelo utiliza embeddings de palabras y posiciones para representar el texto de entrada. BERT se entrena inicialmente en dos tareas principales: el modelado de palabras enmascaradas (Masked Language Modeling) y la predicción de la siguiente oración (Next Sentence Prediction).

3.2.2. BERT-Base-Uncased

El modelo BERT-Base-Uncased es una variante estándar de BERT que no distingue entre mayúsculas y minúsculas en su representación de texto. Tiene las siguientes especificaciones:

- Número de capas: 12.
 - Cabezas de atención: 12.
 - Dimensión oculta: 768.
 - Parámetros totales: 110 millones.
- Tras el preentrenamiento BERT se ajusta para tareas específicas como

clasificación de texto, reconocimiento de entidades o análisis de sentimientos. Este ajuste fino requiere agregar una capa adicional sobre el modelo preentrenado y entrenarlo con un conjunto de datos etiquetado para la tarea objetivo. En la elaboración de este proyecto se usó como entrada los datos de los estudiantes del ITG y se ajustó a los patrones de comportamiento que se logró recopilar. Estudios recientes han demostrado la efectividad de BERT en tareas como el análisis de sentimientos, la clasificación de texto y la predicción de riesgos académicos (Sun et al., 2019).

Los transformers son muy útiles al momento de trabajar con datos no estructurados. Además, la interpretación de los resultados mediante técnicas como SHAP proporciona a los administradores información valiosa sobre los factores subyacentes al éxito o fracaso estudiantil (Lundberg & Lee, 2017). Investigaciones adicionales han explorado el impacto de estas tecnologías en mejorar la personalización del aprendizaje y la retención estudiantil (Zhang et al., 2021).

3.3. Aplicaciones de los modelos Transformer

En la actualidad los modelos Transformers han ido demostrando su alto potencial en diversas tareas de predicción como lo menciona Cholakov (2021) en su investigación en la que se compara dos arquitecturas una estándar RNN utilizando LSTM y una implementación simple de Transformer, ambas arquitecturas con el objetivo de pronosticar como cambiara el precio del índice S&P 500. Se entrenaron con la misma cantidad de datos: el valor de cierre diario del índice desde el 3 de enero de 2000 hasta el 31 de agosto de 2018.

Como resultado se pudo comprobar que la red neuronal recurrente LSTM apenas logra seguir una tendencia, mientras que la arquitectura Transformer es capaz de capturar dependencias más detalladas y utilizarlas para realizar pronósticos futuros. Por ejemplo, a corto plazo, el precio del índice generalmente aumenta después de los informes trimestrales de las grandes empresas en un buen año.

Otra investigación relevante se puede mencionar Alerskans (2021) que plantea generar predicciones del cambio en la temperatura en determinados lugares, busca una alternativa a los modelos anteriormente utilizados para predecir el clima ya que no han tenido resultados suficientemente aceptables. El autor menciona una interesante descripción del modelo Transformer “Para ponerlo de manera simple, el modelo Transformer compara los elementos en una secuencia de entrada, aprende cómo se influyen entre sí y utiliza esto para predecir el siguiente elemento en la secuencia.”, para evaluar el rendimiento del modelo Transformer, se comparó con el pronóstico generado por el GFS, así como con dos modelos de procesamiento estándar: un modelo de regresión lineal y una red neuronal (RN) totalmente conectada y no temporal. La regresión lineal (RL) es un método estándar simple para el procesamiento, y la comparación, mientras que la comparación con la RN evalúa los beneficios de la complejidad adicional que se añade al modelo Transformer, como el uso de autoatención para aprender las dinámicas temporales de los datos.

Otro campo en el que han avanzado los Transformers es el de los Vision Transformers (ViTs), que han emergido como la técnica dominante en varias tareas

de visión y, en muchos casos, han superado a las Redes Neuronales Convolucionales (CNNs), ya que satisfacen el mecanismo de autoatención y capturan relaciones de largo alcance en las imágenes. Según Khawar (2023) la arquitectura de los ViTs se puede lograr en varios pasos clave:

- División de la imagen en parches no superpuestos o superpuestos.
- Aplanado de los parches y generación de embeddings lineales de menor dimensión.
- Adición de embeddings posicionales y un token de clase.
- Alimentación de la secuencia de parches al transformador.
- Paso del valor del token de clase a una capa MLP para obtener la predicción final.

En un estudio realizado por Han, et al. (2023) destaca algunos casos de éxito en los que se puede mencionar que la segmentación de imágenes ha sido uno de ellos. Los ViTs han mostrado resultados sobresalientes. En tareas de segmentación semántica, para instancias y la panóptica. Existen métodos como TransUNet, que ha combinado la alta resolución de las CNNs, junto con la segmentación basada en transformers, para resolver problemas en los que se requiere una tasa de muestreo mucho mayor, de manera tal que se puedan mantener los patrones a un tamaño tal que se puedan identificar como una función de clase, además de redundar en el problema inverso en el que se requieren representaciones de secuencias.

Estos modelos presentan gran efectividad en aplicaciones como: trabajos de traducción automática, creación de textos, análisis de los diferentes

sentimientos emocionales, reconocimiento de voz, entre otros.

También el proceso de predicción con un modelo Transformer puede describirse en las siguientes etapas:

Preprocesamiento de Datos: Antes de realizar el entrenamiento del modelo se realiza el proceso de:

- **Tokenización:** Se divide las secuencias de datos en tokens para que puedan ser procesadas por un modelo de predicción.
- **Codificación Posicional:** Con el fin de mantener el orden lógico en cada token se agrega una posición.

Mecanismo de Atención: El mecanismo de atención permite al modelo enfocarse en las partes más relevantes de la entrada en cada tarea asignada.

Propagación a Través del Modelo: Los resultados del mecanismo de atención se procesan a través de:

- **Capas completamente conectadas:** Ajustan las representaciones generadas por la atención.
- **Normalización:** Regula los valores para evitar la explosión o desaparición de gradientes durante el entrenamiento.

Generación de Salida: El modelo genera predicciones que se basa en las representaciones finales de la secuencia. En tareas de clasificación, la salida puede ser una probabilidad asociada a cada clase.

En resumen, los modelos Transformers han llevado a cabo una gran cantidad de proyectos innovadores en distintos campos de la inteligencia artificial. La sociedad cada vez se vuelve más

dependiente de ellos, ya que son solamente eficaces, flexibles, versátiles, y aparentemente cada vez más capaces de resolver problemas complejos, y a gran escala. Nombrar todos los sistemas que se han construido sobre ellos es muy difícil, pero sí que se pueden mencionar algunos que son realmente interesantes: por ejemplo, GPT4All, Flair, Lama-Cleaner, Recommenders, Sentence-transformers, DeepChem, LaTeX-OCR, Haystack, etc. Esto demuestra el impacto y la flexibilidad de los Transformers, además incentiva la creación de sistemas nuevos para predecir la deserción estudiantil.

4. Resultados

La revisión de la literatura puso de manifiesto que las redes neuronales Transformers han sido utilizadas en diversos ámbitos del aprendizaje automático. Los más relevantes son el procesamiento de lenguaje natural (NLP), la visión por ordenador y la predicción de series temporales. Los modelos de Transformers han mostrado un rendimiento superior en tareas como la traducción automática, la clasificación de texto y la generación de lenguaje en comparación con las redes neuronales recurrentes (RNN), gracias a su capacidad para captar relaciones a largo plazo en los datos y el procesamiento de secuencias en paralelo.

En lo referente a PLN, las investigaciones analizadas en el presente trabajo evidenciaron mejoras significativas de precisión en tareas de detección de entidades nombradas, en clasificación de sentimientos, en generación de resúmenes automáticos entre otras. Los modelos BERT o GPT demostraron tener un buen rendimiento a la hora de ofrecer una buena comprensión contextual del lenguaje,

propiciando avances sobre aspectos como asistentes virtuales o motores de búsqueda o sistemas de recomendación.

En el contexto de la visión por computadora, el llamado Vision Transformers (ViTs) ha mejorado en ciertos aspectos a las arquitecturas de las redes convolucionales (CNN), aplicado para el reconocimiento y la segmentación de varias imágenes. De hecho, se ha llegado a observar que este tipo de modelos son capaces de detectar patrones mucho más complejos, particularmente en el campo biomédico y la seguridad. Hoy en día, ya se encuentran algunas investigaciones sobre la combinación de los Vision Transformers (ViTs) con las redes neuronales convolucionales (CNN) en busca de una mayor eficiencia computacional sin perder rendimiento.

En lo que se refiere a la previsión de series temporales, se identificó que los Transformers se han utilizado en la predicción de datos meteorológicos, financieros y educativos. Investigaciones asociadas a la predicción de la deserción estudiantil han determinado que los Transformes son los modelos que pueden procesar grandes cantidades de datos demográficos y académicos con una habilidad mayor que la de los métodos tradicionales.

A la vez, los artículos revisados mostraron como, los Transformers son modelos muy efectivos, pero también son modelos con retos evidentes, especialmente en la elevada necesidad de consumo computacional y en la necesidad de la disponibilidad de grandes cantidades de datos para poder ser entrenados. También se evidenció en algunas investigaciones recientes que se han encontrado algunas formas de optimizar este tipo de modelos,

por ejemplo: implementando versiones livianas, reduciendo los números de parámetros, o utilizando técnicas de aprendizaje auto-supervisado.

Finalmente, se llegó a identificar que el interés en el uso de Transformers sigue en aumento, tanto en el contexto de la generación de imágenes y de vídeos, como en procesos industriales automatizados, en la localización de anomalías en datos estructurados y no estructurados y en muchos otros ámbitos. Estos resultados son reflejo de la importancia y del alcance de la aplicabilidad de los Transformers en el ámbito de la inteligencia artificial y el aprendizaje profundo, convirtiéndose en una de las tecnologías fundamentales para la evolución de sistemas predictivos y automatizados.

Tabla 1
Aplicaciones, ventajas y desafíos de las redes neuronales Transformers

Área de Aplicación	Ejemplo de Uso	Ventajas	Desafíos
Procesamiento del Lenguaje Natural (PLN)	Traducción automática, generación de texto (GPT, BERT).	Mayor precisión en comprensión de contexto y generación de lenguaje.	Alto costo computacional y necesidad de grandes volúmenes de datos.
Visión por Computadora	Reconocimiento de imágenes y segmentación (Visión Transformers).	Captura de relaciones espaciales y mayor precisión en tareas de visión.	Requiere grandes conjuntos de datos para un rendimiento óptimo.
Predicción de Series Temporales	Predicción financiera, meteorológica y educativa.	Captura de relaciones a largo plazo en datos temporales.	Interpretabilidad limitada y necesidad de técnicas explicativas.
Educación	Predicción de deserción estudiantil.	Identificación temprana de patrones de riesgo académico.	Requiere integración con datos contextuales variados.
Seguridad y Detección de Anomalías	Monitoreo de fraudes y ciberseguridad.	Mayor capacidad para detectar patrones inusuales en grandes volúmenes de datos.	Reducción de falsos positivos y mejora en la interpretabilidad.

Nota. Sintetiza los hallazgos principales del artículo, destacando el impacto de los

Transformers en diferentes áreas y los desafíos que deben abordarse en el futuro. *Fuente:* Elaboración propia

5. Discusión

Los hallazgos de esta revisión avalan que los Transformers han superado a los modelos tradicionales en procesamiento del lenguaje natural (PLN), la visión por computadora y la predicción de series temporales. En el PLN, modelos como BERT y GPT han demostrado una superioridad por sus mayores precisiones en la clasificación de texto o en la generación de lenguaje, sin embargo, perdura un inconveniente que es el alto coste computacional. En la visión por computadora, los Vision Transformers (ViTs) han hecho mejoras en la segmentación y en la detección de imágenes; sin embargo, el rendimiento disminuye al trabajar con pequeños conjuntos de datos.

En el contexto de las predicciones de series temporales, ha de señalarse que la utilización de Transformers para modelar datos de tipo financiero, climatológico o escolar resulta adecuada y eficaz, mostrando relevancia en el área de la deserción estudiantil. Sin embargo, la interpretabilidad del modelo Transformer todavía está bajo investigación, pero se ha propuesto el uso de técnicas de interpretabilidad como SHAP, que proporcionan mayor transparencia.

En las futuras investigaciones se podría plantear la creación de Transformers más eficaces; el diseño de modelos híbridos en el que conjunten Transformers con otras técnicas del aprendizaje automático; y la mejora de la explicabilidad, en particular en el ámbito de la toma de decisiones críticas.

6. Conclusiones

Las redes neuronales han transformado drásticamente muchas áreas de la inteligencia artificial, desde el procesamiento del lenguaje natural hasta la visión por computadora. Con el tiempo, las arquitecturas han ido modificándose desde modelos muy simples como el Perceptrón Multicapa (MLP) o las Redes Neuronales Convolucionales (CNN), hasta las más sofisticadas como las Redes Neuronales Recurrentes (RNN) y, actualmente, las arquitecturas conocidas como Transformers.

Los modelos Transformers han superado a las RNN y LSTM en el tratamiento de datos secuenciales, fundamentalmente gracias a su mecanismo de atención, que tiene la capacidad de capturar dependencias de largo alcance de un modo muy eficiente. Los Transformers, por su capacidad de paralelización y escalabilidad, han llevado a importantes avances en la actualidad, en distintas aplicaciones como la traducción automática, predicción de deserción estudiantil, generación de texto y modelos de lenguaje a gran escala como es el caso de BERT y GPT.

En concreto, en el ámbito educativo y el empresarial, los modelos de aprendizaje automático han contribuido a tomar decisiones con base en datos, haciendo posible la personalización del aprendizaje o la optimización de estrategias de retención estudiantil. En los campos financiero y meteorológico, los Transformers han demostrado ser mejores que los modelos tradicionales en aquellas tareas de predicciones de tendencias y el análisis de patrones complejos.

Sin embargo, como todo tiene sus ventajas e inconvenientes, los Transformers presentan ciertas limitaciones como el alto coste computacional y a la necesidad de

muchas muestras de datos durante el entrenamiento de estos. Por otro lado, aspectos como la interpretabilidad de los modelos o la ética en el uso de la inteligencia artificial siguen siendo puntos controversiales dentro de la comunidad científica.

Con el tiempo, la evolución de las redes neuronales podría estar definida por la combinación y adición de modelos híbridos, la computación cuántica, y los métodos más eficientes energéticamente. La unión de Transformers con métodos de optimización y modelos explicativos puede permitir soluciones más adecuadas y accesibles en diferentes áreas aplicativas.

En definitiva, las redes neuronales continúan hoy en día en constante desarrollo, mostrando nuevas posibilidades en el área de la automatización, de la inteligencia artificial y del procesamiento de la información. Su influencia sobre la sociedad seguirá ampliándose, convirtiéndose en una de las tecnologías primordiales para llevar a cabo la transformación digital del siglo XXI.

7. Desafíos a futuro

A pesar de la versatilidad y éxito de las redes neuronales Transformers en un conjunto considerable de aplicaciones, hay muchos problemas por resolver para que su uso sea óptimo, y además para incrementar su accesibilidad.

Alto costo computacional y consumo de recursos

Los modelados de tipo Transformers necesitan grandes volúmenes de datos y de potentes unidades de procesamiento que pueden ser GPU o TPU para obtener los mejores resultados, lo que provoca dudas acerca de la adopción de este tipo de

modelo en situaciones con recursos limitados. Futuras investigaciones deben centrarse en arquitecturas más eficientes, como Transformers ligeros para reducir el número de parámetros sin comprometer su rendimiento.

Falta de transparencia e interpretabilidad

La toma de decisiones fundamentada en modelos de aprendizaje profundo constituye un problema sin resolver, por la falta de explicabilidad inherente de los Transformers. SHAP y LIME han sido usados para incrementar la transparencia, pero se necesitan más investigaciones sobre metodologías que reduzcan la incertidumbre en la interpretación de sus predicciones sobre todo en aquellas aplicaciones que son críticas (salud, educación, finanzas).

Optimización del entrenamiento en conjuntos de datos pequeños

Si bien los modelos Transformer han demostrado su eficacia en el manejo de grandes volúmenes de datos, su comportamiento puede no ser tan robusto en situaciones donde hay pocos datos disponibles. Desarrollar técnicas de aprendizaje auto-supervisado, transferencia de aprendizaje y preentrenamiento eficiente podría ayudar a hacer que sean más aplicables a situaciones en las que hay pocos datos disponibles o en las que conseguir datos resulta muy costoso.

Eficiencia en tareas en tiempo real

Una gran variedad de aplicaciones, entre ellas el análisis de vídeo, la traducción en tiempo real y sistemas de recomendación, requieren modelos que devuelvan respuestas rápidas. La reducción de la latencia y la mejora de la eficiencia

computacional de los Transformers serán un pilar básico para su utilización en aplicaciones interactivas y en aquellos sistemas con restricciones temporal.

Seguridad y sesgo algorítmico

Los Transformers pueden generar sesgos latentes que se encuentran presentes en los datos de entrenamiento, y producir resultados discriminatorios o que no sean correctos. Es por ello que se hace necesario crear diferentes estrategias para la detección y la mitigación de sesgos para que sean más justos y equitativos en la toma de decisiones.

Expansión a nuevos dominios

Si bien los Transformers han demostrado ser exitosos en PLN y visión por computadora, su potencial en otras áreas como la robótica, la predicción de eventos climáticos extremos y el modelado de sistemas biológicos todavía está en camino de exploración. Adaptar estas arquitecturas a nuevos dominios podría abrir nuevas oportunidades para su aplicación.

Abordar estos desafíos permitirá que los Transformers sean más accesibles, eficientes y confiables, consolidando su papel en la evolución de la inteligencia artificial y el aprendizaje profundo.

Contribución de los autores: Los autores han contribuido en todos los apartados de la investigación.

Financiamiento: Los autores financiaron totalmente el estudio.

Conflictos de intereses: Los autores declaran no tener ningún conflicto de intereses.

Referencias

- Alerskans, E., Nyborg, J., Birk, M., & Kaas, E. (2022). *A transformer neural network for predicting near-surface temperature*. Recuperado el 9 de enero de 2025 de <https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/met.2098>
- Agrusti F, B. G. (2020). *Deep learning approach for predicting university dropout: A case study at roma tre university*. *Journal of E-Learning and Knowledge Society*. Recuperado el 7 de febrero de 2025 de <https://bit.ly/3tlirGE>
- Briceño, M. (2024). *Análisis comparativo de algoritmos de Machine Learning y Deep Learning para la predicción de la Deserción Estudiantil Universitaria*. Recuperado el 9 de enero de 2025 de <https://rmets.onlinelibrary.wiley.com/doi/epdf/10.1002/met.2098>
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning phrase representations using RNN encoder-decoder for statistical machine translation (arXiv:1406.1078)*. *arXiv*. Recuperado el 9 de enero de 2026 de <https://arxiv.org/pdf/1406.1078>
- Cholakov, R. & Kolev, T. (2021). *Transformers predicting the future. applying attention in next-frame and time series forecasting*. Recuperado el 15 de enero de 2025 de <https://arxiv.org/pdf/2108.08224>

- Campos, S. & Pastor O. (2023). *Implementación de un método de clasificación para detectar la deserción de estudiantes de la carrera de Ingeniería de Industrias Alimentarias de una Universidad Nacional Peruana basado en aprendizaje de máquina*. Recuperado el 5 de febrero de 2025 de <https://repositorio.uss.edu.pe/handle/20.500.12802/12096>
- Coraspe, R. (2020). *Student Dropout Prediction*. Recuperado el 4 de enero de 2025 de <https://github.com/Wittline/Dropout-Students-Prediction>
- Dávila, G. (2023). *Modelo de aprendizaje automático para la predicción de deserción de estudiantes de la carrera de Informática de la UTPL*. Recuperado el 3 de febrero de 2025 de https://dspace.utpl.edu.ec/visorHub/?handle=123456789_51902
- Devlin, J., Chang, M.-W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of deep bidirectional transformers for language understanding*. Recuperado el 10 de enero de 2026 de <https://arxiv.org/pdf/1810.04805>
- De la Torre, J. (2023). *Transformadores: Fundamentos teóricos y aplicaciones*. Recuperado el 3 de julio de 2024 de <https://arxiv.org/abs/2302.09327>
- González, F. (2005). *Repitencia y Deserción Universitaria en América Latina*. Informe sobre a Educación Superior en América Latina y el Caribe: UNESCO-IESALC. Recuperado el 7 de julio de <https://unesdoc.unesco.org/ark:/48223/pf0000140087>
- Han, K. (2023). *A Survey on Visual Transformer*. Recuperado el 8 de julio de 2024 de <https://arxiv.org/pdf/2012.12556>
- Haro, J. (2023). *Modelo de aprendizaje profundo para la predicción de deserción de estudiantes de la carrera de informática*. Recuperado el 2 de febrero de 2025 de https://dspace.utpl.edu.ec/visorHub/?handle=123456789_51939
- Hochreiter, S., & Schmidhuber, J. (1997). *Long short-term memory*. *Neural Computation*, 9(8), 1735–1780. Recuperado el 10 de enero de 2026 de <https://doi.org/10.1162/neco.1997.9.8.1735>
- Himmel, E. (2002). *Modelo de análisis de la deserción estudiantil en la educación superior*. Recuperado el 2 de febrero de 2025 de <https://dialnet.unirioja.es/descarga/articulo/8310798.pdf>
- Khawar, I. (2023). *Recent Advances in Vision Transformer: A Survey and Outlook of Recent Work*. Recuperado el 8 de julio de 2024 de 2203.01536 (arxiv.org)
- LeCun, Y., Bengio, Y., & Hinton, G. (2015). *Deep Learning*. Recuperado el 8 de enero de 2025 de <https://www.researchgate.net/profile/Y->

- Bengio/publication/277411157_Deep_Learning/links/55e0cdf908ae2fac471ccf0f/Deep-Learning.pdf?_tp=eyJjb250ZXh0ljp7ImZpcnN0UGFnZSI6InB1YmXPY2F0aW9uIn19
- Lee, Y., Shin, D., Loh, H., Lee, J., Chae, P., Cho, J., Park, S., Lee, J., Baek, J., Kim, B., & Choi, Y. (2020). *Deep Attentive Study Session Dropout Prediction in Mobile Learning Environment*. Recuperado el 3 de julio de 2024 de <https://arxiv.org/pdf/2002.11624.pdf>
- Lee, S. & Young, J. (2019). *The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction*. Recuperado el 1 de febrero de 2025 de <https://www.mdpi.com/2076-3417/9/15/3093/pdf?version=1564561987>
- Aulck, L., Velagapudi, N., Blumenstock, J., & West, J. (2016). *Predicting Student Dropout in Higher Education*. Recuperado el 5 de febrero de 2025 de <https://arxiv.org/abs/1606.06364>
- Molina, A. (2023). *Predicting Student Dropout in Higher Education*. Recuperado el 5 de julio de 2024 de <https://repositorio.uniandes.edu.co/server/api/core/bitstreams/c51474d1-6999-4392-ab87-176360818fe1/content>
- Montalvo, J. Sánchez, N. (2023). *Modelo de predicción de deserción: Un estudio de caso de estudiantes de la Universidad Politécnica Estatal del*
- Carchi*. Recuperado el 5 de febrero de 2025 de <https://prometeojournal.com.ar/index.php/prometeo/article/view/34/53>
- Perez, R., Rodríguez, A., Durán, D., Reverón, C. & Pinto, M. (2007). *Cuestión de supervivencia. Graduación, deserción y rezago en la Universidad Nacional de Colombia*. Bogotá: Universidad Nacional de Colombia. Revista teoría y praxis investigativa. Recuperado el 8 de julio de 2024 de <https://dialnet.unirioja.es/descarga/articulo/3063193.pdf>
- Ramírez, M. (2025). *Predicción de la Deserción Estudiantil Universitaria utilizando Deep Learning. Caso: Instituto Superior Tecnológico Quito*. Recuperado el 16 de marzo de 2025.
- Song, Z., Sung, S., Park, D., & Park, B. (2023). *All-Year Dropout Prediction Modeling and Analysis for University Students*. *Applied Sciences*. Recuperado el 2 de julio de 2024 de <https://www.mdpi.com/2076-3417/13/2/1143>
- Sun, H. & Joon, S. (2021). *Early Dropout Prediction in Online Learning of University using Machine Learning*. Recuperado el 2 de febrero de 2025 de <https://bit.ly/3WR2IAN>
- Sunbok, L., & Jae, Y. (2019). *The Machine Learning-Based Dropout Early Warning System for Improving the Performance of Dropout Prediction*. Recuperado el 2 de febrero de 2025 de <https://bit.ly/3tlirGE>

- Teleamazonas. (s.f.). *Estudio de la SENESCYT sobre deserción universitaria en Ecuador*. Recuperado el 1 de julio de 2024 de <https://www.teleamazonas.com/estudio-senescyt-desercion-universitaria-ecuador/#:~:text=Seg%C3%BAn%20un%20estudio%20de%20indicadores,hombres%20que%20en%20las%20mujeres.>
- Tinto, V. (2012). *Definir la deserción: Una cuestión de perspectiva*. Revista de Educación Superior. Recuperado el 9 de julio de 2024 de http://publicaciones.anuies.mx/pdfs/revista/Revista71_S1A3ES.pdf
- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gómez, A. N., Kaiser, Ł., & Polosukhin, I. (2017). *Attention is all you need*. In *Advances in neural information processing systems*. Recuperado el 3 de julio de 2024 de <https://arxiv.org/pdf/1706.03762>
- Villar, A. & Velini, C. (2023). *Supervised machine learning algorithms for predicting student dropout and academic success: a comparative study*. *Discover Artificial Intelligence*. Recuperado el 1 de julio de 2024 de <https://link.springer.com/article/10.1007/s44163-023-00079-z>
- Won, H., Kim, M., Kim, D., Kim, H., & Kim, K. (2023). *University Student Dropout Prediction Using Pretrained Language Models*. Recuperado el 2 de julio de 2024 de <https://www.mdpi.com/2076-3417/13/12/7073>
- Zambrano, G., Rodríguez, K., & Guevara, L. (2018). *Análisis de la deserción estudiantil en las universidades del Ecuador y América Latina*. Revista pertinencia académica. Recuperado el 9 de julio de 2024 de <https://revistas.utb.edu.ec/index.php/rpa/article/view/2451/2059>