Generación de noticias a partir de conjuntos de datos económicos utilizando técnicas de inteligencia artificial y aprendizaje automático

David Leonardo López Gómez
Luis Vidal Bejarano Bejarano

Proyecto Aplicado para optar al título de
Magister en Ciencia de Datos

Director(a)

Diego Luis Linares Ospina

Codirector(a)

Gloria Inés Álvarez Vargas

FACULTAD DE INGENIERÍA Y CIENCIAS
MAESTRÍA EN CIENCIA DE DATOS
ENERO DE 2025

Tabla de contenido

Resumen 7

Introducción 1

1. Definición del problema 2

1.1. Planteamiento del problema 2

2. Objetivos del proyecto 6

2.1. Objetivo general 6

2.2. Objetivos específicos 6

3. Antecedentes y marco teórico 7

3.1. Antecedentes 7

3.2. Marco Teórico 10

3.2.1. Procesos ETL en la transformación de datos económicos 10
3.2.2. Modelos de lenguaje natural para la generación de noticias 11
3.2.3. Aprendizaje automático para la generación de contenido informativo 12
3.2.4. APIs como herramienta para la integración de datos en tiempo real 13

3.3. Metodologías 14

3.3.1. Modelos Transformers 14
3.3.2. Arquitectura de los modelos transformers 17
3.3.3. Modelos GPT-2 21
3.3.4. Modelo T5 23
Etapas de Implementación del Modelo T5 26
3.3.5. Métricas de desempeño 27

4. Pre-procesamiento y procesamiento de datos 31

4.1. Datos 31

4.1.1. Construcción del set de datos 31

4.2. Limpieza y pre-procesamiento del texto 33

5. Caracterización, optimización de parámetros y construcción de modelos de redacción de noticias 35

5.1. Entrenamiento de los modelos 35

5.2. Modelo T5 37

5.3. Modelo GPT2 39

5.4. Optimización de parámetros 41

5.4.1. Optimización de parámetros modelo T5 43
5.4.2. Resultados con base en la optimización de parámetros T5 46
5.4.3. Optimización de parámetros modelo GPT2 48
5.4.4. Resultados con base en la optimización de parámetros GPT2 50

5.5. Evaluación del desempeño en la generación de noticias 52

6. Redacción de textos, criterios de comparación y evaluación de los modelos implementados 54

6.1. Redacción modelo T5 sin optimizar 54

6.2. Redacción modelo GPT2 sin optimizar 57

6.3. Redacción final modelos T5 y GPT2 con parámetros optimizados 61

7. Conclusiones 65

8. Referencias Bibliográficas 67

LISTA DE FIGURAS

Gráfica 1. Arquitectura de modelos transformers 18
Gráfica 2. Proceso de búsqueda bayesiana con Optuna 44

LISTA DE TABLAS

Tabla 1. Metodologías de generación de textos 9
Tabla 2. Componentes de la métrica BLEU 27
Tabla 3. Componentes de la métrica BERTScore 28
Tabla 4. Listado de variables de la base de datos 32
Tabla 5. Número de observaciones por temática 33
Tabla 6. Proceso de limpieza de datos 34
Tabla 7. Listado de hiperparámetros iniciales del modelo T5 41
Tabla 8. Métricas versión inicial o básica del modelo T5 42
Tabla 9. Listado de hiperparámetros iniciales del modelo GPT2 42
Tabla 10. Métricas versión inicial o básica del modelo GPT2 43
Tabla 11. Hiperparámetros evaluados con búsqueda bayesiana con Optuna 45
Tabla 12. Parámetros de temperature y top_p evaluados con la grilla 45
Tabla 13. Métricas con optimización de parámetros - modelo T5 47
Tabla 14. Hiperparámetros evaluados con búsqueda bayesiana con Optuna 49
Tabla 15. Parámetros de temperature y top_p evaluados con la grilla 50
Tabla 16. Métricas con optimización de parámetros - modelo GPT2 51
Tabla 17. Noticias redactadas con el modelo T5 54
Tabla 18. Noticias redactadas con el modelo GPT2 57
Tabla 19. Noticias redactadas con los modelos T5 y GPT2 con parámetros optimizados 61

Resumen

En el contexto actual de rápida generación de datos económicos, surge el desafío de transformar conjuntos de datos complejos en información accesible y comprensible. Esta investigación aborda esta problemática mediante el uso de técnicas de inteligencia artificial (IA) y aprendizaje automático (AA) para la generación automatizada de noticias basadas en procesamiento de lenguaje natural (PLN). Aquí, el problema radica en la ineficiencia de los métodos tradicionales de análisis de datos frente al volumen y la complejidad de la información económica contemporánea. Además, los medios de comunicación enfrentan una creciente demanda de entrega rápida y precisa de contenidos, lo que presenta desafíos significativos en términos de costos y operatividad.

Para abordar estos retos, se desarrolló un sistema basado en los modelos T5 y GPT-2, ambos con arquitecturas de tipo transformer. Los datos económicos fueron preprocesados, limpiados y estructurados para entrenar los modelos. Se realizaron ajustes utilizando hiperparámetros optimizados y métricas de evaluación como BLEU, BERTScore y perplejidad, con el objetivo de medir la calidad de los textos generados. Los resultados mostraron que el modelo T5 superó al GPT-2 en precisión, coherencia y fluidez del texto generado, logrando un BLEU de 0.14, un BERTScore promedio (F1) de 0.83 y una perplejidad de 1.11 tras la optimización. En comparación, el GPT-2 alcanzó un BLEU de 0.15, un BERTScore promedio (F1) de 0.70 y una perplejidad de 12.75. Estas métricas indican que el T5 es más adecuado para generar textos complejos y altamente estructurados, mientras que el GPT-2 destaca en tareas donde se requiere mayor creatividad y generación de contenido más diversificado.

El resultado de esta investigación, aporta un avance en la transformación de datos económicos en contenido informativo, reduciendo costos y tiempos asociados a los procesos tradicionales. La solución propuesta se presenta como una herramienta prometedora para democratizar el acceso a información económica y apoyar la toma de decisiones en entornos dinámicos.

Introducción

En el contexto actual, la información económica es un recurso fundamental para comprender, analizar y prever el comportamiento de los mercados financieros y el desarrollo económico. Sin embargo, el volumen exponencial de datos económicos representa tanto un desafío como una oportunidad. Transformar estos datos en información accesible y comprensible es crucial, especialmente en un entorno digital competitivo donde el periodismo enfrenta presiones para implementar nuevas tecnologías que impulsen modelos de negocio más sostenibles, mejoren la eficiencia y ofrezcan servicios de valor agregado.

El avance de la inteligencia artificial (IA) y el aprendizaje automático (AA) ha abierto nuevas posibilidades en la generación automatizada de noticias, con un enfoque particular en el procesamiento de lenguaje natural (PLN). Estas tecnologías permiten analizar grandes volúmenes de datos económicos, identificar patrones, tendencias y eventos clave, y convertir esta información en narrativas precisas y contextualizadas. De esta forma, la IA transforma la cadena de valor del periodismo, desde la recopilación de datos hasta la producción, distribución y consumo de noticias.

El presente proyecto propone el desarrollo de un sistema que utilice modelos avanzados de PLN para la generación automatizada de noticias a partir de conjuntos de datos económicos. Estos modelos no solo facilitarán el acceso a información económica clave, sino que también transformarán datos complejos en contenidos claros, relevantes y fáciles de comprender para audiencias diversas.

Los resultados esperados incluyen la creación de un modelo capaz de generar noticias económicas de manera coherente y contextual, con base en un conjunto de datos. Asimismo, la implementación de algoritmos eficientes de PLN mejorará la calidad y la precisión de las noticias generadas, asegurando que estas sean oportunas y fiables. En última instancia, esta solución busca no solo democratizar el acceso a información económica compleja, sino también impulsar una toma de decisiones más informada en un entorno global cada vez más interconectado.

1. Definición del problema

1.1. Planteamiento del problema

Debido a la creciente velocidad con la que se generan datos e información económica en los últimos años, las técnicas tradicionales de análisis han derivado en baja productividad y uso ineficiente de los recursos [1]. [2], expresa en su investigación que, el volumen de datos al que se enfrentan los periodistas es tan grande que resulta imposible encontrar una noticia si no se diseña una aplicación que permita agrupar y analizar variables. Asimismo, [3] y [4] identificaron que en las últimas dos décadas, el aumento exponencial de información en medios electrónicos obligó a los diferentes medios de comunicación a comprender, de manera rápida, el contenido relevante de estos grandes volúmenes de información textual.

Por otra parte, según el Fondo Monetario Internacional, los datos económicos se han convertido en un insumo crítico de la producción económica moderna, junto con la tierra, el capital, la mano de obra y el petróleo. Así las cosas, el valor de los datos económicos han generado la necesidad de construir textos fundamentados en el procesamiento de lenguaje natural, el cual, permite construir textos (noticias) basados en información de entrada [5], [6].

En relación con lo anterior, la Asociación de Medios de Información – AMI, indica que la construcción de noticias automatizadas asegura que los periódicos pueden utilizar estas técnicas y así generar grandes volúmenes de contenido rápido, preciso y confiable. Asimismo, la BBC expresa que el periodismo económico automatizado ayuda a los redactores a reducir tiempos y costos de oportunidad, a lo cual se enfrentan los diarios si no publican sus noticias con mayor velocidad, consistencia y precisión.

Una vez abordados los diferentes enfoques, se presenta la perspectiva desde la ciencia de datos, dado que existen diferentes retos para la construcción de noticias a partir de conjuntos de datos económicos. De acuerdo con diferentes autores [4], [6]–[9], los principales retos se pueden asociar

con:

Construcción de procesos: la fragmentación o partición de los procesos puede afectar el flujo de trabajo y replicabilidad del modelo. Así las cosas, es obligatorio el conocimiento de los datos y del "negocio" para la construcción de los procesos.

Estandarización de los datos: dado el componente de automatización del proceso, se hace necesario estandarizar los datos que alimentarán el modelo, dado que, si esto no se cumple, puede que los resultados del modelo se vean afectados. Adicionalmente, es necesario revisar los procesos de alistamiento y preparación de la información.

Modelación: este punto se resume en la selección de la técnica de modelado, que permita construir el caso de prueba y el modelo. Aquí se pueden combinar diferentes técnicas de minería de datos. En esta etapa de construcción del modelo, se deben establecer parámetros específicos que serán utilizados en la etapa de evaluación.

Evaluación: de acuerdo con diferentes parámetros de los modelos desarrollados, se podrá evaluar la bondad de ajuste, la precisión y demás características que permitan elegir el modelo que más se ajuste a los requerimientos solicitados.

Desarrollo: aquí se desarrollará el modelo y la guía de uso del mismo. Además, se incluirán las rutinas de software necesaria para su aplicabilidad. Es importante mencionar que este punto no puede ser desarrollado sin haber superado los retos previos.

Para el caso concreto de construir noticias a partir de conjuntos de datos económicos. Los retos se resumirían en:

Una fase de entendimiento de los datos económicos. Identificar la coherencia de los mismos y comprender el sentido de los mismos a la hora de transmitir un mensaje. Por ejemplo: un aumento de la tasa de desempleo es un dato negativo, pero un aumento del PIB es positivo. Así las cosas, los aumentos y reducciones en los indicadores deben incluir una naturaleza que esté inmersa en el modelo.

Preparación de los datos. Aquí se debe construir una ruta para el alistamiento de la información, dado que esta será el insumo principal del modelo que permitirá escribir las noticias. Es importante

mencionar que la presente investigación se desarrollará con datos abiertos (o libres) y que responden a variables generales de la economía y que requieren ser etiquetados para la modelación.
Modelación. Una vez se hace el alistamiento de los datos, se debe construir el caso de prueba y el modelo que responda a la necesidad. En este punto se podrán combinar diferentes técnicas de minería de datos y obtener diversos modelos.
Evaluación. Aquí se deberán evaluar los diferentes modelos y seleccionar el que mejor se ajuste a los datos. Así las cosas, el mejor modelo tendrá que brindar buenas medidas de ajuste estadístico y mayor claridad a la hora de transmitir las noticias y/o mensajes.
Desarrollo. Construir el modelo y la guía de uso del mismo. Además, se deberán incluir las rutinas de software necesarias para su aplicación.

En este sentido, el problema se plantea en términos de generar noticias a partir de datos económicos, utilizando técnicas de inteligencia artificial y aprendizaje automático.

Formulación del problema

Para poder abordar el problema que se plantea, se proponen los siguientes interrogantes:

¿Cómo se realizará el pre-procesamiento y procesamiento de datos y/o cifras económicas, para que estos sirvan de insumo para un modelo que escriba noticias a partir de estos?
Dado que se utilizarán datos y/o cifras económicas, que requieren de una lógica para entender su sentido ¿cuál estrategia se aplicará para el entrenamiento de los modelos que escribirán las noticias con base en estos datos?
¿Cómo se extraerán las características relevantes para los o el modelo seleccionado, dado que el resultado final es una notica redactada?
¿Cuál será el criterio de evaluación de los modelos seleccionados, dado que el resultado final es una notica redactada?

Pregunta general:

¿Cómo desarrollar un modelo para la generación de noticias a partir de conjuntos de datos económicos, que utilice técnicas de inteligencia artificial y aprendizaje automático?

2. Objetivos del proyecto

2.1. Objetivo general

Generar noticias a partir de conjuntos de datos económicos, utilizando técnicas de inteligencia artificial y aprendizaje automático.

2.2. Objetivos específicos

- Establecer una estrategia de pre-procesamiento y procesamiento de datos y/o cifras económicas que permitan alimentar un modelo que brinde como resultado una noticia de perfil económico sustentada en dichos insumos.
- Extraer las características relevantes o con mayor potencial que le permitan al modelo generar una escritura inicial de las noticias.
- Construir una estrategia de entrenamiento aplicable a los modelos implementados, que permita al modelo tener insumos para la interpretación y comprensión del lenguaje a la hora de construir las noticias.
- Desarrollar uno o varios modelos que permitan construir noticias a partir de conjuntos de datos económicos.
- Aplicar un criterio de evaluación a los modelos seleccionados.

3. Antecedentes y marco teórico

3.1. Antecedentes

Durante los últimos años, los diferentes periódicos y medios de comunicación han sufrido una profunda renovación basada en el continuo progreso de la tecnología, en particular la inteligencia artificial [3]. En relación con lo anterior, [5] indica que, en la actualidad se cuenta con gran cantidad de información en documentos de texto digitales, y cuando un usuario está interesado en profundizar en una cierta temática, esta puede estar contenida en gran cantidad de textos relacionados que difícilmente pueden revisados en su totalidad, teniendo así que invertir tiempo y esfuerzo para encontrar de manera fácil lo que está buscando. Asimismo, [3] y [4] identificaron que en las últimas dos décadas, el aumento exponencial de información en medios electrónicos hace necesario comprender, de manera rápida, el contenido relevante de estos grandes volúmenes de información textual.

Dado este contexto, es importante resaltar que, la construcción automática de textos es una labor fundamentada en el procesamiento de lenguaje natural, el cual, tiene como objetivo construir textos que se alimentan de una entrada de información predeterminada [5], [6]. Es importante mencionar que, el procesamiento de lenguaje natural (PLN) tiene sus orígenes en la inteligencia artificial, allí se quería dotar a las computadoras de un sistema de inteligencia capaz de procesar el lenguaje natural para entender, reproducir, inferir y deducir el conocimiento presente en la información [4], [6].

Asimismo, [4] proponen que "*la generación automática de texto (GART) es una de las tareas prioritarias del área de investigación de procesamiento de lenguaje natural"*, con esto se pretende construir textos similares a los desarrollados por humanos, pero con ahorros significativos de recursos. Según [7], los modelos de generación de texto están revolucionando el ámbito al generar texto similar al humano. Así, han ganado gran popularidad recientemente en muchos ámbitos como noticias, redes sociales, escritura de guiones de películas y composición de poesía, entre otros.

Es así como, la capacidad de descifrar una cantidad tan enorme de datos para extraer información útil es una tarea importante y requiere un mecanismo automático para ayudar con el depósito de información existente [10]. De acuerdo con [11], la construcción de texto a partir de datos puede ser generalmente categorizada según el tipo de entrada del modelo (simple o múltiple), el propósito (genérico, específico o basado en consultas) y tipo de salida (extractiva o abstractiva).

En relación con lo anterior, existen diferentes enfoques para la construcción de textos, sin embargo, tres de estos resaltan por su aceptación en términos de investigación y son: enfoque basado en frecuencia, enfoque basado en características y enfoque basado en el aprendizaje automático [11]. En relación con esto último, entra el concepto de aprendizaje automático, que de acuerdo con [10] es utilizar un conjunto de datos de entrenamiento para entrenar un modelo, una vez hecho esto, se construyen modelos de probabilidad que permiten elegir una frase u otra según los insumos. Algunos de los métodos comunes de aprendizaje automático son: método de Naive Bayes, método de red neuronal artificial y método de lógica difusa. Asimismo, [11] indica que el objetivo del aprendizaje automático puede estar estrechamente relacionado con un problema de clasificación, es decir, el modelo debe aprender de un entrenamiento para determinar la clase adecuada a la que pertenece un elemento de los que está observando.

Por otro lado, autores como [7], expresan que la generación de texto es un campo de estudio en el procesamiento del lenguaje natural (PNL), el cual combina la lingüística computacional y la inteligencia artificial para generar texto nuevo. En este sentido, es un proceso de generación de texto sintético gramatical y semánticamente correcto. El cual, incluye entrenar un modelo que toma datos de entrada, enseñarle el contexto de la entrada y generar un texto nuevo relacionado con el dominio de los datos utilizados. Y, por último, el texto generado debe satisfacer la estructura básica del lenguaje y transmitir el mensaje deseado

En resumen, la estrategia básica de generación de texto es primero entrenar cualquier modelo de lenguaje en muchas secuencias de datos de texto, y luego el modelo es capaz de generar el siguiente carácter o varios caracteres en la secuencia dados los caracteres anteriores como entrada [4]–[8].

En cuanto a las diferentes estrategias de generación de textos con base en bases de datos, diferentes autores [12]–[14] identifican como fundamentales las siguientes metodologías (ver tabla 1).

Tabla 1. Metodologías de generación de textos

Metodología	Modelo	Características
Redes neuronales recurrentes (RNN)	Redes neuronales recurrentes (RNN)	Memoria interna (problema de gradiente que desaparece)
	Redes de memoria a corto plazo (LSTM)	Puede descubrir dependencias a largo plazo.
Redes generativas de confrontación (GAN)	Redes generativas de confrontación (GAN)	Entrenamiento adverso
	SeqGAN	GAN's recuenciales
	LeakGAN	Permite abordar el problema de la generación de textos largos.
Codificadores automáticos variacionales (VAE)	Codificadores automáticos variacionales (VAE)	Puede operar con datos sin etiquetar
	Skip-VAE	Permite enfrentar el problema del "colapso de variable latente"
Modelos generados con base en transformaciones	GPT2 – Modelos transformadores generativos pre-entrenados	Bloque de-codificador y representaciones unidireccionales.
	BERT – Modelos de representaciones de codificación bidireccional de transformadores	Bloque codificador y representaciones bidireccionales

Elaboración propia con base en [12]–[14].

Por último, surge el interrogante de la evaluación del desempeño del modelo, aquí [10] habla de la posibilidad de construir métricas intrínsecas para evaluar el rendimiento de un texto automatizado comparándolo frente a un texto "ideal". Allí mismo, los autores hablan del término de precisión, recordación y desempeño de los textos construidos a partir de datos.

La estructura del documento contempla cinco secciones. En la primera de ellas, se plantea una

revisión de literatura que permita construir el contexto de la investigación. Posteriormente, se presentarán las diferentes metodologías de construcción automática de textos. En la tercera parte, se hará la construcción del modelo de generación automática de noticias (textos) económicas. En la cuarta, evaluaremos el resultado del modelo y la construcción de la noticia "ideal". Por último, concluiremos.

3.2. Marco Teórico

En la actual economía del conocimiento, los datos han emergido como el recurso más valioso para la toma de decisiones y la creación de contenidos informativos. Según [10], mientras que las revoluciones industriales previas se fundamentaron en recursos físicos como la máquina de vapor, la electricidad y el petróleo, la revolución actual se basa en los datos, un insumo etéreo pero omnipresente que sustenta la transformación digital.

La capacidad de recolectar, almacenar y analizar datos masivos provenientes de dispositivos móviles, sensores, Internet de las Cosas (IoT) y big data ha abierto nuevas oportunidades para la generación automatizada de contenido. En el contexto del periodismo económico, esta capacidad tecnológica permite transformar conjuntos de datos económicos en noticias relevantes y comprensibles, democratizando el acceso a información clave para audiencias diversas.

En este proceso, es esencial distinguir entre la calidad de los datos y la calidad de la información. La calidad de los datos hace referencia a las características intrínsecas de la materia prima, mientras que la calidad de la información evalúa si el producto final satisface las necesidades del usuario [12]. Este principio guía la generación automatizada de noticias económicas, asegurando que las narrativas producidas sean precisas, relevantes y útiles.

3.2.1. Procesos ETL en la transformación de datos económicos

Para generar noticias económicas a partir de datos, es fundamental preparar la información

mediante procesos ETL (Extract, Transform, Load) [13]. Estas etapas permiten la integración de datos provenientes de diversas fuentes y su transformación en insumos listos para análisis y generación automatizada de contenido.

Extracción
La extracción implica recolectar datos económicos de sistemas de origen, como bases de datos financieras, APIs de instituciones económicas y otras fuentes públicas o privadas. Este paso incluye:
- Validar la estructura y formato de los datos.
- Convertirlos al formato requerido para su análisis posterior.

Transformación
En esta fase, los datos extraídos se procesan y normalizan aplicando reglas específicas que aseguren su coherencia y relevancia. Estas transformaciones permiten identificar patrones, tendencias y eventos clave que serán integrados en las noticias generadas.

Carga
Los datos transformados se almacenan en sistemas de destino, como bases de datos o data warehouses, desde donde pueden ser utilizados para alimentar modelos de aprendizaje automático y de lenguaje natural que automatizan la generación de contenido.

3.2.2. Modelos de lenguaje natural para la generación de noticias

Los modelos de procesamiento de lenguaje natural (NLP, por sus siglas en inglés) han revolucionado la forma en que las máquinas interactúan con el lenguaje humano. Entre estos, los modelos basados en arquitecturas de Transformers, como *GPT (Generative Pre-trained Transformer)* y *T5 (Text-to-Text Transfer Transformer)*, han demostrado capacidades sobresalientes para tareas como generación de texto, traducción automática, resumen de textos y clasificación de información.

GPT (Generative Pre-trained Transformer)
GPT, desarrollado por OpenAI, es un modelo de lenguaje autoregresivo que utiliza una arquitectura

Transformer para predecir la siguiente palabra en una secuencia de texto. Su entrenamiento en grandes cantidades de datos textuales le permite generar contenido coherente y contextualizado, adaptándose a una amplia variedad de temas y estilos. En el contexto de la generación de noticias económicas, GPT puede:

- Redactar narrativas económicas claras y estructuradas.
- Identificar relaciones entre eventos económicos para destacar implicaciones clave.
- Adaptar el tono y estilo de las noticias según la audiencia objetivo, desde expertos financieros hasta el público general.

T5 (Text-to-Text Transfer Transformer)

T5, desarrollado por Google, adopta un enfoque text-to-text, donde todas las tareas de NLP, como clasificación, traducción o generación de texto, se convierten en problemas de generación de texto. Este enfoque unificado lo hace altamente flexible y efectivo en la generación de contenido narrativo. Para la generación de noticias económicas, T5 puede:

- Resumir grandes volúmenes de datos económicos en textos concisos.
- Generar titulares atractivos y contextualmente relevantes.
- Realizar análisis comparativos basados en datos históricos y actuales.

Ambos modelos pueden ser complementados con técnicas específicas de ajuste fino (fine-tuning) en conjuntos de datos económicos, lo que mejora su capacidad para generar narrativas precisas y adaptadas a contextos económicos.

3.2.3. Aprendizaje automático para la generación de contenido informativo

El aprendizaje automático (Machine Learning) complementa el uso de modelos de lenguaje natural al permitir la extracción y análisis de patrones en los datos económicos. Según [14], este enfoque dota a las máquinas de la capacidad de aprender y generalizar conocimientos a partir de datos. En el caso de la generación de noticias económicas, los modelos de aprendizaje supervisado y no supervisado pueden:

- Identificar eventos económicos relevantes.

- Detectar anomalías en tendencias económicas.
- Proveer datos estructurados que sirvan como insumos para los modelos de lenguaje natural.

Al integrar técnicas de aprendizaje automático con modelos de lenguaje como GPT o T5, se crea un flujo de trabajo que combina análisis profundo de datos con narrativas claras y coherentes.

3.2.4. APIs como herramienta para la integración de datos en tiempo real

Las APIs (Application Programming Interfaces) juegan un papel crucial en la automatización de la generación de noticias económicas al facilitar el acceso a datos en tiempo real. Según [15], una API es un conjunto de reglas que permite la comunicación entre aplicaciones, actuando como puente para integrar datos externos en sistemas automatizados.

En el ámbito económico, las APIs de bancos centrales, bolsas de valores o plataformas de análisis financiero permiten acceder a información actualizada, como:
- Indicadores macroeconómicos (PIB, inflación, tasas de interés).
- Datos de mercados financieros (precios de acciones, tipos de cambio).
- Informes de instituciones globales (FMI, Banco Mundial).

Estas fuentes de datos en tiempo real alimentan los modelos de lenguaje natural, permitiéndoles generar contenido dinámico y contextualizado en función de los eventos económicos más recientes. En resumen, este marco teórico integra los fundamentos necesarios para la generación automatizada de noticias económicas utilizando técnicas de inteligencia artificial y aprendizaje automático. Destaca cómo los procesos ETL, los modelos avanzados de lenguaje como GPT y T5, y las APIs en tiempo real trabajan en conjunto para transformar datos económicos complejos en narrativas claras y accesibles. Esta combinación no solo mejora la eficiencia en la producción de contenido, sino que también democratiza el acceso a información económica clave, facilitando la toma de decisiones informadas en un entorno global dinámico.

3.3.Metodologías

Esta sección de metodologías describe las herramientas y enfoques seleccionados para dar cumplimiento al objetivo de *"Generar noticias a partir de conjuntos de datos económicos, utilizando técnicas de inteligencia artificial y aprendizaje automático"*. Dada la naturaleza compleja y dinámica de los datos económicos, se optó por utilizar modelos avanzados de procesamiento de lenguaje natural (NLP), específicamente GPT-2 y T5, ambos basados en arquitecturas de Transformers. Estos modelos han demostrado capacidades excepcionales para generar textos coherentes, relevantes y contextualmente adaptados, características esenciales para la producción de contenido informativo de alta calidad.

Tanto GPT-2, desarrollado por OpenAI, como T5, creado por Google, son ampliamente reconocidos por su capacidad para procesar y generar lenguaje humano con una notable precisión. Estas herramientas permitirán transformar datos económicos en narrativas informativas, claras y accesibles. Mientras GPT-2 se centra en la generación de texto autoregresivo, ideal para elaborar contenidos continuos y estilísticamente coherentes, T5 utiliza un enfoque text-to-text que facilita tareas como resúmenes, creación de titulares y redacción de noticias adaptadas a distintos contextos.

3.3.1. Modelos Transformers

De acuerdo con [15], los transformers son un tipo de arquitectura de red neuronal que transforma o cambia una secuencia de entrada en una secuencia de salida. Para ello, aprenden el contexto y rastrean las relaciones entre los componentes de la secuencia. Por ejemplo, considere esta secuencia de entrada: *"¿De qué color es el cielo?"*. El modelo transformador usa una representación matemática interna que identifica la relevancia y la relación entre las palabras color, cielo y azul. Usa esa información para generar el resultado: *"El cielo es azul"*.

Las organizaciones usan modelos de transformers para todo tipo de conversiones de secuencias, desde el reconocimiento de voz hasta la traducción automática y el análisis de secuencias de

proteínas.

¿Cómo funcionan los transformers?

Las redes neuronales han sido el método líder en varias tareas de la IA, como el reconocimiento de imágenes y la NLP, desde principios de la década del 2000. Consisten en capas de nodos informáticos interconectados, o neuronas, que imitan al cerebro humano y trabajan juntos para resolver problemas complejos.

Las redes neuronales tradicionales que se ocupan de secuencias de datos suelen utilizar un patrón de arquitectura de codificador/decodificador. El codificador lee y procesa toda la secuencia de datos de entrada, como una oración en inglés, y la transforma en una representación matemática compacta. Esta representación es un resumen que captura la esencia de la entrada. Luego, el decodificador toma este resumen y, paso a paso, genera la secuencia de salida, que podría ser la misma oración traducida al francés.

Este proceso ocurre de forma secuencial, lo que significa que tiene que procesar cada palabra o parte de los datos una tras otra. El proceso es lento y puede perder algunos detalles más finos en largas distancias.

Por otro lado, los modelos de transformers modifican este proceso al incorporar algo llamado mecanismo de auto-atención. En lugar de procesar los datos en orden, el mecanismo le permite al modelo observar diferentes partes de la secuencia a la vez y determinar qué partes son las más importantes.

Imagine que está en una habitación concurrida e intenta escuchar a alguien hablar. El cerebro se centra automáticamente en su voz mientras desconecta los ruidos menos importantes. La auto-atención permite al modelo hacer algo similar: presta más atención a los fragmentos de información relevantes y los combina para hacer mejores predicciones de resultados. Este mecanismo hace que los transformers sean más eficientes, lo que les permite entrenarse en conjuntos de datos más

grandes. También es más eficaz, especialmente cuando se trata de textos largos en los que el contexto lejano puede influir en el significado de lo que viene después.

Tipos de modelos de transformers

Transformers bidireccionales

Las representaciones de codificadores bidireccionales de los modelos de transformers (BERT) modifican la arquitectura base para procesar las palabras en relación con todas las demás palabras de una oración y no de forma aislada. Técnicamente, emplea un mecanismo llamado modelo de lenguaje enmascarado bidireccional (MLM). Durante el pre-entrenamiento, las BERT ocultan aleatoriamente un porcentaje de los tokens de entrada y predicen estos tokens ocultos basándose en su contexto. El aspecto bidireccional proviene del hecho de que las BERT tienen en cuenta las secuencias de tokens de izquierda a derecha y de derecha a izquierda en ambas capas para una mayor comprensión.

Transformers generativos pre-entrenados

Los modelos de transformers generativos pre-entrenados (GPT) utilizan decodificadores de transformers apilados que se entrenan previamente en un gran corpus de texto mediante el uso de objetivos de modelado del lenguaje. Son autorregresivos, lo que significa que retroceden o predicen el siguiente valor de una secuencia en función de todos los valores anteriores. Al utilizar más de 175 000 millones de parámetros, los modelos de GPT pueden generar secuencias de texto que se ajustan según el estilo y el tono. Los modelos de GPT han impulsado la investigación en IA para lograr la inteligencia artificial general. Esto significa que las organizaciones pueden alcanzar nuevos niveles de productividad y, al mismo tiempo, reinventar sus aplicaciones y experiencias de los clientes.

Transformers bidireccionales y autorregresivos

Un transformador bidireccional y autorregresivo (BART) es un tipo de modelo de transformador

que combina propiedades bidireccionales y autorregresivas. Es como una mezcla del codificador bidireccional de BERT y el decodificador autorregresivo de GPT. Lee toda la secuencia de entrada a la vez y es bidireccional como BERT. Sin embargo, genera la secuencia de salida de un token a la vez, condicionada a los tokens generados previamente y a la entrada proporcionada por el codificador.

Transformers para tareas multimodales

Los modelos de transformers multimodales, como ViLBERT y VisualBERT, están diseñados para administrar varios tipos de datos de entrada, normalmente texto e imágenes. Amplían la arquitectura del transformador mediante el uso de redes de doble flujo que procesan entradas visuales y textuales por separado antes de fusionar la información. Este diseño permite que el modelo aprenda las representaciones intermodales. Por ejemplo, ViLBERT usa capas transformadoras coatencionales para permitir que los flujos separados interactúen. Es crucial para situaciones en las que es clave comprender la relación entre el texto y las imágenes, como las tareas visuales de respuesta a preguntas.

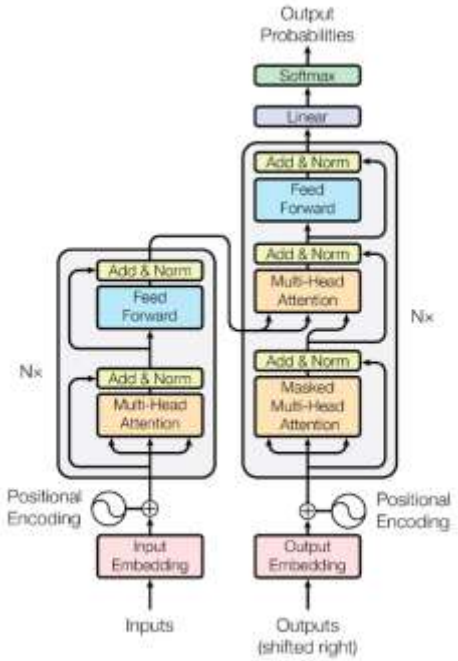
Transformers de visión

Los transformers de visión (ViT) reutilizan la arquitectura del transformador para tareas de clasificación de imágenes. En lugar de procesar una imagen como una cuadrícula de píxeles, ven los datos de la imagen como una secuencia de parches de tamaño fijo, de forma similar a como se tratan las palabras en una oración. Cada parche se aplana, se incrusta linealmente y luego se procesa secuencialmente mediante el codificador de transformador estándar. Se agregan incrustaciones posicionales para mantener la información espacial. Este uso de la auto-atención global permite al modelo capturar las relaciones entre cualquier par de parches, independientemente de su posición.

3.3.2. Arquitectura de los modelos transformers

La arquitectura de la red neuronal del transformador tiene varias capas de software que trabajan juntas para generar el resultado final.

Gráfica 1. Arquitectura de modelos transformers



Fuente: https://aws.amazon.com/es/what-is/transformers-in-artificial-intelligence/

Incrustaciones de entrada

Esta etapa convierte la secuencia de entrada en el dominio matemático que entienden los algoritmos de software. Al principio, la secuencia de entrada se divide en una serie de tokens o componentes de secuencia individuales. Por ejemplo, si la entrada es una oración, los tokens son palabras. La incrustación transforma entonces la secuencia del token en una secuencia vectorial matemática. Los vectores contienen información semántica y sintáctica, representada como números, y sus atributos se aprenden durante el proceso de entrenamiento.

Puede visualizar los vectores como una serie de coordenadas en un espacio n-dimensional. Como ejemplo simple, piense en una gráfica bidimensional, donde x representa el valor alfanumérico de la primera letra de la palabra e y representa sus categorías. La palabra banana tiene el valor (2,2) porque comienza con la letra b y está en la categoría fruta. La palabra mango tiene el valor (13,2)

porque comienza con la letra m y también está en la categoría fruta. De esta forma, el vector (x, y) le dice a la red neuronal que las palabras banana y mango pertenecen a la misma categoría.

Ahora imagine un espacio n-dimensional con miles de atributos sobre la gramática, el significado y el uso de cualquier palabra en oraciones asignadas a una serie de números. El software puede usar los números para calcular las relaciones entre las palabras en términos matemáticos y comprender el modelo del lenguaje humano. Las incrustaciones permiten representar los tokens discretos como vectores continuos que el modelo puede procesar y de los que puede aprender.

Codificación posicional

La codificación posicional es un componente crucial en la arquitectura del transformador porque el modelo en sí no procesa inherentemente los datos secuenciales en orden. El transformador necesita una forma de considerar el orden de las fichas en la secuencia de entrada. La codificación posicional agrega información a la incrustación de cada token para indicar su posición en la secuencia. Esto se hace a menudo mediante el uso de un conjunto de funciones que generan una señal posicional única que se agrega a la incrustación de cada token. Con la codificación posicional, el modelo puede preservar el orden de los símbolos y comprender el contexto de la secuencia.

Bloque transformador

Un modelo de transformador típico tiene varios bloques de transformers apilados juntos. Cada bloque transformador tiene dos componentes principales: un mecanismo de auto-atención con múltiples cabezales y una red neuronal de retroalimentación por posición. El mecanismo de auto-atención permite al modelo sopesar la importancia de las diferentes fichas dentro de la secuencia. Al hacer predicciones, se centra en las partes relevantes de la entrada.

Por ejemplo, piense en las frases “Speak no lies (No diga mentiras)” y “He lies down (Se acuesta)”. En ambas oraciones, el significado de la palabra lies no se puede entender sin mirar las palabras

que están al lado. Las palabras speak y down son fundamentales para entender el significado correcto. La auto-atención permite agrupar los elementos relevantes para el contexto.

La capa de alimentación directa tiene componentes adicionales que ayudan al modelo del transformador a entrenarse y funcionar de manera más eficiente. Por ejemplo, cada bloque transformador incluye lo siguiente:

- Conexiones en torno a los dos componentes principales que actúan como accesos directos. Permiten el flujo de información de una parte de la red a otra, omitiendo ciertas operaciones intermedias.
- La normalización de capas mantiene los números (específicamente las salidas de las diferentes capas de la red) dentro de un rango determinado para que el modelo se capacite sin problemas.
- La transformación lineal funciona para que el modelo ajuste los valores para realizar mejor la tarea en la que se está entrenando, como el resumen del documento en lugar de la traducción.

Bloques lineales y softmax

En última instancia, el modelo necesita hacer una predicción concreta, como elegir la siguiente palabra de una secuencia. Aquí es donde entra en juego el bloque lineal. Es otra capa totalmente conectada, también conocida como capa densa, antes de la etapa final. Realiza un trazado lineal aprendido desde el espacio vectorial hasta el dominio de entrada original. Esta capa crucial es donde la parte del modelo para la toma de decisiones toma las complejas representaciones internas y las convierte de nuevo en predicciones específicas que puede interpretar y utilizar. El resultado de esta capa es un conjunto de puntuaciones (a menudo denominadas logits) para cada token posible.

La función softmax es la etapa final que toma las puntuaciones logit y las normaliza en una distribución de probabilidad. Cada elemento de la salida de softmax representa la confianza del modelo en una clase o token en particular.

3.3.3. Modelos GPT-2

El modelo GPT-2 (Generative Pre-trained Transformer 2), desarrollado por OpenAI[1], es una de las arquitecturas más influyentes en el ámbito del procesamiento de lenguaje natural (NLP). Su diseño se basa en la arquitectura de Transformers, introducida por [16], que utiliza mecanismos de atención para procesar y generar texto de manera eficiente y precisa. GPT-2 es un modelo de lenguaje autoregresivo, lo que significa que genera texto palabra por palabra, prediciendo cada palabra en función de las anteriores dentro del contexto.

El modelo GPT-2 se basa en una arquitectura de Transformer que utiliza mecanismos de atención y representaciones de texto basadas en el modelo de lenguaje auto-regresivo. A continuación, se presenta un esquema de ecuaciones que describe su funcionamiento:

- *Representación de entrada:* el texto de entrada se convierte en una secuencia de tokens y luego se transforma en vectores de embeddings:

$$\boldsymbol{x} = \{x_1, x_2, \dots, x_n\}, \quad \mathbf{E} = \text{Embedding}(x_i)$$

Donde:

$x_i = Tokens\ del\ texto$

$\mathbf{E} = Embedding\ de\ dimensión\ fija\ para\ cada\ token.$

- *Positional Encoding:* dado que los Transformers no tienen una estructura recurrente, se agrega información posicional a los embeddings:

$$\mathbf{E}' = \mathbf{E} + \mathbf{P}$$

Donde:

$\mathbf{P} =$ Codificación posicional, definida generalmente como una combinación de funciones sinusoidales.

- *Multi-Head Attention:* el mecanismo de atención es la base del modelo. Para cada cabeza de

[1] OpenAI es una empresa estadounidense de investigación y despliegue de inteligencia artificial fundada en 2015 e, inicialmente, sin ánimo de lucro. Su misión original era asegurar que la inteligencia artificial general fuese desarrollada como código libre y en beneficio de toda la humanidad.

atención, se calculan las siguientes matrices:

Proyección de las entradas en queries, keys y values:

$$\boldsymbol{Q} = \boldsymbol{E}'\boldsymbol{W}_Q, \qquad \boldsymbol{K} = \boldsymbol{E}'\boldsymbol{W}_K, \qquad \boldsymbol{V} = \boldsymbol{E}'\boldsymbol{W}_V$$

Cálculo de la atención escalada:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\text{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}$$

Combinar las cabezas de atención:

$$\text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{Concat}(head_1, head_2, \dots, head_h)\boldsymbol{W}_O$$

Donde:

$\boldsymbol{W}_Q, \boldsymbol{W}_K, \boldsymbol{W}_V, \boldsymbol{W}_O$: *Pesos aprendidos del modelo.*

d_k: *Dimensión de cada cabeza de atención.*

- *Capa de Normalización y Residuos:* el resultado de la atención multi-cabeza se normaliza y se combina con la entrada original:

$$\boldsymbol{Z}_1 = LayerNorm\big(\mathbf{E}' + \text{MultiHead}(\mathbf{Q}, \mathbf{K}, \mathbf{V})\big)$$

- *Capa de Red Neuronal Feed-Forward:* el modelo utiliza una red totalmente conectada con activación no lineal para procesar la salida:

$$\boldsymbol{Z}_2 = LayerNorm\big(\boldsymbol{Z}_1 + \text{FFN}(\boldsymbol{Z}_1)\big)$$

Donde:

$$\text{FFN}(\boldsymbol{z}) = ReLU(\boldsymbol{z}\boldsymbol{W}_1 + \boldsymbol{b}_1)\boldsymbol{W}_2 + \boldsymbol{b}_2$$

- *Función de Pérdida (Autoregresiva):* GPT-2 predice el siguiente token $x_t + 1$ dado el contexto previo $\{x_1, x_2, \dots, x_t\}$ minimizando la entropía cruzada:

$$\mathcal{L} = -\sum_{t=1}^{n} \log P(x_{t+1}|x_1, x_2, \dots, x_t)$$

Donde:

$$P(x_{t+1}|x_1, x_2, \dots, x_t) = \text{softmax}\big(\mathbf{Z}_t\mathbf{W}_{output} + \mathbf{b}_{output}\big)$$

- *Generación de Texto (Decodificación):* durante la generación, el modelo predice token por token mediante un enfoque auto-regresivo:

$$x_{t+1} = \text{argmax}(\log P(x_{t+1}|x_1, x_2, \dots, x_t))$$

Estas etapas se pueden resumir es:

- Entrada y Embedding: tokenización y representación en vectores.

- Codificación Posicional: agrega contexto secuencial.

- Multi-Head Attention: calcula relaciones entre palabras.

- Normalización y Residuos: estabiliza la entrada a la siguiente capa.

- Red Feed-Forward: procesa cada palabra en un espacio de características más abstracto.

- Salida: predicción del próximo token basada en probabilidades.

Etapas de Implementación de GPT-2

- Recolección de Datos: Se seleccionan textos económicos de noticias económicas, para entrenar el modelo.
- Pre-procesamiento: Los datos se normalizan para asegurar consistencia en formato, estilo y terminología.
- Ajuste Fino: El modelo GPT-2 se entrena nuevamente en este corpus específico para personalizar su desempeño en el dominio económico.
- Evaluación y Validación: Se evalúa la calidad del texto generado en términos de precisión, coherencia y relevancia contextual, comparándolo con narrativas redactadas por humanos.

En resumen, GPT-2 representa una herramienta poderosa para automatizar la generación de noticias económicas, proporcionando una solución eficiente y escalable para convertir datos complejos en narrativas claras y accesibles. Su metodología se basa en aprovechar su capacidad de comprensión contextual y generación lingüística, adaptándolo específicamente a las necesidades del proyecto.

3.3.4. Modelo T5

El modelo T5 (Text-to-Text Transfer Transformer), desarrollado por Google Research[2], es un

[2] Es una comunidad global de investigadores que impulsan avances científicos con Google Cloud.

enfoque innovador que redefine la forma de abordar tareas de procesamiento de lenguaje natural (NLP). A diferencia de los modelos tradicionales, T5 convierte todas las tareas de NLP en problemas de entrada y salida textuales. Este diseño simplifica el entrenamiento y la implementación, al tratar tareas como clasificación, traducción, resumen y generación de texto dentro de un mismo marco de trabajo.

- *Entrada del modelo:* el modelo T5 recibe una secuencia de texto como entrada, representada como un conjunto de tokens $\boldsymbol{X} = \{x_1, x_2, \dots, x_n\}$. La entrada es tokenizada utilizando un tokenizador pre-entrenado, donde cada token x_i es mapeado a un vector de dimensión d, representando la palabra en un espacio vectorial.

$$\boldsymbol{X} = Tokenize(texto_{entrada})$$

Donde:

$\boldsymbol{X} \in \mathbb{R}^{nxd}$ es la secuencia de tokens transformada en vectores (matriz de entrada).

- *Codificador (Encoder):* el codificador toma la secuencia de entrada y la mapea a una representación interna $h_{encoder}$. Este proceso se realiza mediante una serie de capas de transformadores.

$$h_{encoder} = TransformerEncoder(X)$$

Donde:

$h_{encoder} \in \mathbb{R}^{nxd_{hidden}}$ es la representación interna del texto de entrada.

- *Multi-Head Attention:* la atención en el modelo T5 se realiza a través de la Multi-Head Attention, donde las representaciones del codificador se ajustan para capturar dependencias globales. La operación de atención se puede escribir como:

$$\text{Attention}(\mathbf{Q}, \mathbf{K}, \mathbf{V}) = \text{softmax}\left(\frac{\boldsymbol{Q}\boldsymbol{K}^{\mathrm{T}}}{\sqrt{d_k}}\right)\boldsymbol{V}$$

Donde:

$Q \in \mathbb{R}^{nxd_Q}$ es la consulta (query)

$K \in \mathbb{R}^{nxd_K}$ es la clave (key)

$V \in \mathbb{R}^{nxd_V}$ es el valor (value)

d_k es la dimensión de la clave

Esta operación se aplica en varias capas del codificador y decodificador.

- *Decodificador (Decoder):* el decodificador toma las representaciones del codificador y genera la secuencia de salida. En este caso, la salida es una secuencia de tokens $\boldsymbol{Y} = \{y_1, y_2, \dots, y_m\}$ que representa la noticia generada.

$$\boldsymbol{Y} = \text{TransformerDecoder}(h_{encoder})$$

Donde:

$\boldsymbol{Y} \in \mathbb{R}^{mxd}$ es la secuencia generada de tokens (noticia).

- *Generación de Secuencia:* la salida del decodificador se pasa por una capa de proyección lineal para obtener las puntuaciones de probabilidad para cada token del vocabulario.

$$\hat{Y} = \text{softmax}(\boldsymbol{W}_{out}\boldsymbol{h}_{decoder} + \boldsymbol{b}_{out})$$

Donde:

$\widehat{\boldsymbol{Y}} \in \mathbb{R}^{mxV}$ es la distribución de probabilidad de los tokens en el vocabulario.

$\boldsymbol{W}_{out} \in \mathbb{R}^{Vxd_{hidden}}$ es una matriz de pesos de salida

$\boldsymbol{b}_{out} \in \mathbb{R}^{V}$ es un vector de sesgo

$\boldsymbol{V}$ es el tamaño del vocabulario

- *Cálculo de la Pérdida (Loss):* la función de pérdida utilizada en T5 es típicamente la entropía cruzada (cross-entropy), que mide la discrepancia entre la distribución de probabilidad generada $\widehat{\boldsymbol{Y}}$ y las etiquetas reales $\boldsymbol{Y}_{true}$.

$$\mathcal{L} = -\sum_{t=1}^{m} Y_{true,t} \,.\log\left(\widehat{\boldsymbol{Y}}_t\right)$$

Donde:

$Y_{true,t}$ es el token verdadero en la posición t

$\widehat{\boldsymbol{Y}}_t$ es la probabilidad del token en la posición t

- *Optimización:* la optimización del modelo T5 se realiza utilizando un algoritmo como Adam para

minimizar la función de pérdida:

$$\theta^* = arg\min_{\theta} \mathcal{L}(\theta)$$

Donde:

θ representa los parámetros del modelo (pesos y sesgos)

$\mathcal{L}(\theta)$ es la función de perdida

El modelo T5, en su arquitectura de codificador-decodedor, utiliza un proceso de atención para capturar las dependencias entre las palabras de la entrada y la salida. A través de múltiples capas de transformadores, el modelo genera una secuencia de tokens de salida a partir de una entrada de datos económicos. El modelo es entrenado mediante la minimización de la pérdida de entropía cruzada, utilizando el algoritmo de optimización Adam.

Etapas de Implementación del Modelo T5

- Recolección de Datos: se recopilan textos económicos, en su totalidad noticias económicas. Los datos se limpian para garantizar la inclusión de información precisa y actualizada, eliminando redundancias y errores.
- Pre-procesamiento de Datos: los textos recopilados se normalizan, etiquetan y estructuran siguiendo el formato text-to-text. Esto incluye la creación de pares entrada-salida, en este caso noticias y cifras económicas.
- Fine-Tuning del Modelo T5: el modelo T5 se ajusta utilizando los datos pre-procesados. Durante este proceso, el modelo aprende a transformar datos estructurados en narrativas informativas, especializándose en el lenguaje y estilo característicos del periodismo económico.
- Generación de Noticias: T5 se utiliza para convertir entradas como indicadores económicos o análisis de tendencias en noticias claras y comprensibles.
- Validación y Evaluación: se evalúa la calidad del texto generado en aspectos como precisión, relevancia y coherencia. Las narrativas se comparan con noticias redactadas por periodistas humanos para garantizar que cumplan con estándares editoriales.

En conclusión, el modelo T5 destaca como una herramienta poderosa y versátil para abordar la generación automatizada de noticias económicas. Su enfoque text-to-text permite una adaptación eficiente a tareas complejas, como convertir datos económicos estructurados en narrativas comprensibles y contextualmente relevantes. El uso de T5 en este proyecto no solo asegura precisión y coherencia en el texto generado, sino que también ofrece flexibilidad para adaptarse a múltiples formatos narrativos y estilos editoriales.

3.3.5. Métricas de desempeño

A través de un enfoque sistemático, y con la premisa de cumplir con el objetivo general de la investigación, en el desarrollo de los modelos se evaluarán combinaciones de parámetros, midiendo su impacto en métricas de desempeño como BLEU, BERTScore y perplejidad. La descripción de estas métricas es la siguiente:

BLEU (Bilingual Evaluation Understudy)

Es una métrica común para evaluar la calidad de texto generado automáticamente, como traducciones o resúmenes, comparándolos con referencias humanas. La estructura del cálculo se presenta a continuación:

Tabla 2. Componentes de la métrica BLEU

Precisión a nivel de n-grama	Penalización por longitud	Cálculo del BLEU
El BLEU calcula la precisión de los n-gramas generados frente a los de referencia. $$\boldsymbol{P_n} = \frac{\sum_{c \in C} \boldsymbol{clip}(\boldsymbol{n-gramas_c})}{\sum_{c \in C} \boldsymbol{n-gramas_c}}$$ Donde: C: Conjunto de todas las oraciones generadas. $\boldsymbol{n-gramas_c}$: Total de n-gramas en la oración generada c	Se introduce para evitar que el modelo genere textos muy cortos solo para maximizar la precisión de los n-gramas. $$BP = \begin{cases} 1 & si\ l_c > l_r \\ e^{1-\frac{l_r}{l_c}} & si\ l_c \leq l_r \end{cases}$$ Donde: l_c: Longitud total de las oraciones generadas.	El BLEU se define como el promedio geométrico ponderado de las precisiones a nivel de n-gramas, multiplicado por la penalización por longitud: $$BLEU = BP.\, e^{\left(\sum_{n=1}^{N} \omega_n \log P_n\right)}$$ Donde: ω_n: Peso asignado a cada n-grama

$\boldsymbol{clip(n-gramas_c)}$: Cuenta máxima de coincidencias de n-gramas generados con las referencias.	l_r: Longitud total de las oraciones de referencia más cercanas.

Fuente: elaboración propia 2024.

BERTScore

Esta métrica evalúa la calidad de texto generado automáticamente. A diferencia de BLEU, que se basa en coincidencias exactas de n-gramas, BERTScore utiliza representaciones vectoriales obtenidas de modelos de lenguaje como BERT para comparar la similitud semántica entre el texto generado y el texto de referencia. A continuación, se presenta el funcionamiento de la métrica.

Tabla 3. Componentes de la métrica BERTScore

Etapa	Descripción
Embeddings semánticos	Para cada palabra (o token) en el texto generado y el texto de referencia, se extraen representaciones vectoriales (embeddings) utilizando un modelo pre-entrenado como BERT
Cálculo de similitud entre tokens	Se utiliza la similitud del coseno para medir la relación entre los embeddings de las palabras generadas y las palabras de referencia: $$\text{sim}(u,v) = \frac{u.v}{\|u\|\|v\|}$$ donde u y v son los vectores de los embeddings de dos palabras
Emparejamiento óptimo de palabras	Para cada palabra en el texto generado, se encuentra la palabra más parecida en el texto de referencia (y viceversa) según la similitud del coseno. Esto asegura que las palabras sean comparadas con las más relevantes.
Cálculo de las métricas	Precisión: La precisión mide como los tokens generados están alineados semánticamente con los tokens de referencia: $$P = \frac{1}{\|T_g\|}\sum_{t_g \in T_g} max_{t_r \in T_r}\,\text{sim}(t_g, t_r)$$ Donde: $\|T_g\|$: Número total de tokens en el texto generado.

$\text{sim}(t_g, t_r)$: Similitud del coseno entre el embedding del token t_g (generado) y t_r (referencia).

$max_{t_r \in T_r}$: Encuentra el token de referencia t_r con la mayor similitud con t_g

Recuperación:

La recuperación mide qué tan bien los tokens de referencia están alineados semánticamente con los tokens generados:

$$P = \frac{1}{|T_r|} \sum_{t_r \in T_r} max_{t_g \in T_g} \text{sim}(t_r, t_g)$$

Donde:

$|T_r|$: Número total de tokens en el texto referencia.

$\text{sim}(t_r, t_g)$: Similitud del coseno entre el embedding del token t_r (referencia) y t_g (generado).

$max_{t_g \in T_g}$: Encuentra el token de generado t_g con la mayor similitud con t_r

F1-Score:

El F1-Score combina la precisión y la recuperación en una única métrica armónica para equilibrar ambos aspectos:

$$F1 = \frac{2.P.R}{P+R}$$

Fuente: elaboración propia 2024.

Perplejidad

Es una métrica utilizada principalmente para evaluar modelos de lenguaje. Mide que tan acertado es un modelo al predecir un conjunto de datos (o una secuencia de texto) y se relaciona directamente con la probabilidad de las palabras predichas por el modelo. A continuación, se presenta el funcionamiento de la métrica:

Definición: La perplejidad se define como la raíz geométrica inversa de la probabilidad del texto generada por el modelo.

$$PPL = 2^{-\frac{1}{N}\sum_{i=1}^{N} log_2 P(\omega_i | \omega_1, \omega_2, \ldots, \omega_{i-1})}$$

En términos de logaritmo natural sería:

$$PPL = \exp(-\frac{1}{N}\sum_{i=1}^{N} log\, P(\omega_i|\omega_1, \omega_2, \dots, \omega_{i-1}))$$

Donde:

N: Número total de palabras (tokens) en el conjunto de datos o secuencia evaluada.

$P(\omega_i|\omega_1, \omega_2, \dots, \omega_{i-1})$: Probabilidad asignada por el modelo al token ω_i dado su contexto previo $\omega_1, \omega_2, \dots, \omega_{i-1}$.

$-\frac{1}{N}\sum_{i=1}^{N} log\, P(\omega_i|\omega_1, \omega_2, \dots, \omega_{i-1})$: Promedio negativo de los logaritmos de las probabilidades, conocido como entropía cruzada.

Entropía cruzada

La entropía cruzada es una métrica fundamental utilizada en el entrenamiento de modelos como T5 y GPT-2, sirve para medir la calidad de las predicciones que realizan durante la generación de texto. La entropía cruzada mide la discrepancia entre dos distribuciones de probabilidad:

- Distribución real (objetivo): Representa las palabras o tokens reales en el texto de referencia, generalmente como una distribución de probabilidad en la que el token correcto tiene una probabilidad de 1 (distribución de una sola categoría).
- Distribución predicha: Generada por el modelo, asigna probabilidades a todos los tokens posibles del vocabulario en cada paso de la generación.

La fórmula de la entropía cruzada es la siguiente:

$$H(p, q) = -\sum_{i=1}^{N} p(y_i).\log\big(q(y_i)\big)$$

Donde:

$p(y_i)$: probabilidad real del token y_i en el texto objetivo.

$q(y_i)$: probabilidad asignada por el modelo al token y_i

N: número total de tokens en la secuencia.

4. Pre-procesamiento y procesamiento de datos

En el desarrollo de un modelo para la generación de noticias económicas, el manejo adecuado de los datos es un paso crucial para garantizar la calidad y relevancia del contenido generado. El pre-procesamiento y procesamiento de datos constituyen las bases para transformar conjuntos de datos económicos crudos en información estructurada y útil. Estas etapas son esenciales para eliminar inconsistencias, seleccionar características clave, y organizar la información de manera que pueda ser eficientemente interpretada por modelos de aprendizaje automático, como T5 y GPT-2. A continuación, se describen el procedimiento.

4.1. Datos

Como todo proyecto en ciencia de datos, estos requieren de diferentes técnicas de pre-procesamiento y procesamiento, las cuales buscan garantizar un análisis eficiente, preciso y significativo de la información. Con esta premisa, el proyecto de *"Generación de noticias a partir de conjuntos de datos económicos utilizando técnicas de inteligencia artificial y aprendizaje automático"*, inició con un requerimiento de información en texto, explícitamente noticias económicas escritas en español. Para cumplir con lo anterior, se revisaron diferentes compendios de datos e información que pudieran convertirse en el insumo principal del proyecto, y al final se decidió crear un set de datos para el proyecto.

4.1.1. Construcción del set de datos

La construcción del set de datos para este proyecto se llevó a cabo con el objetivo de garantizar una representación adecuada y diversificada de la información económica. Para ello, se recopiló contenido textual proveniente de diversas fuentes confiables, abarcando un amplio rango de eventos y análisis económicos en Colombia. En particular, se extrajeron datos de periódicos como El Tiempo, El Espectador, Portafolio, Semana, y La República, entre otros. Estas publicaciones

fueron seleccionadas debido a su alta cobertura de noticias económicas. Las características del data set son las siguientes:

Variables

El set de datos está compuesto por las siguientes variables, las cuales responden a cada uno de los componentes de las noticias utilizadas en el proyecto:

Tabla 4. Listado de variables de la base de datos

Nombre de la variable	Descripción
Año	Indica el año en que se desarrolla la noticia.
Mes	Indica el mes, en caso de qué el texto lo utilice, en que se desarrolla la noticia.
Trimestre	Indica el trimestre, en caso de qué el texto lo utilice, en que se desarrolla la noticia. Esta categoría se observa principalmente en noticias relacionadas con el PIB o la balanza de pagos, estadísticas que publican resultados con esa frecuencia.
Frecuencia	Esta variable puedo tomar valores de mensual, trimestral o anual, según como lo indique la noticia.
Año_comparación	Esta variable indica el año con el cual se está haciendo la comparación de resultados, si y sólo sí, la noticia lo indica.
País	Esta variable indica el país del cual se está hablando en la noticia.
Tendencia	Indica la dirección del hecho económico de la noticia, los valores pueden ser crecimiento o disminución.
Sector	Esta variable solo tiene valores de economía, es una constante en toda la base de datos.
Indicador	Esta variable identifica las temáticas de las noticias que componen la base de datos, que pueden ser: exportaciones, importaciones, inflación, inversión extranjera directa, PIB (producto interno bruto), precio del petróleo y tasa desempleo.
Valor_actual	Esta es la cifra principal de la noticia.
Variación	Aquí se incluye la variación que la cifra actual tuvo con relación con otro periodo de tiempo, si y sólo si, el texto de la noticia así lo indica.
Valor_comparación	Esta es la cifra del periodo anterior contra el cual se compara la cifra del "valor actual".
Entidad	Aquí se incluye el nombre de la entidad que genera la estadística.
Titular	Aquí se incluye el texto del titular de la noticia.
Noticia	Aquí se incluye el texto de la noticia.

Fuente: elaboración propia 2024.

El listado de variables que se describieron en la tabla 2, pueden entenderse en el modelo como insumos de entrada, todas las variables que están dentro del texto de la noticia, y una salida, el texto de la noticia económica.

$$\underbrace{Año, Mes, \dots, Valor_{comparación}, Entidad}_{entrada} = \underbrace{noticia\ económica}_{salida}$$

Tamaño del set de datos

El tamaño de la base de datos que se utilizó para el desarrollo del proyecto fue de 3.188 observaciones, las cuales se distribuyeron de la siguiente forma por temática, que responden dentro de la base de datos a la variable "indicador":

Tabla 5. Número de observaciones por temática

Temática "indicador"	Número de observaciones
Exportaciones	383
Importaciones	358
Inflación	456
Inversión extranjera directa	521
PIB	579
Precio del petróleo	303
Tasa de desempleo	588
Total general	**3.188**

Fuente: elaboración propia 2024.

Es importante indicar que se trabajó en dejar las diferentes temáticas comparables en número de observaciones.

4.2. Limpieza y pre-procesamiento del texto

El proceso de limpieza y pre-procesamiento del texto es una fase fundamental en la construcción de modelos generativos basados en aprendizaje profundo. Este paso garantiza que los datos sean comprensibles y estructurados para el modelo, mejorando así la calidad de las predicciones y la coherencia de los textos generados. Con base en lo anterior, los datos de texto que se obtuvieron para el proyecto recibieron el siguiente tratamiento y para ejemplificar este proceso, se describe por medio de pasos del proceso:

Tabla 6. Proceso de limpieza de datos

Etapa	Descripción
1	Inicialmente, se realiza la carga de los datos, los cuales contienen información numérica y textual. Para facilitar la interpretación del modelo, los valores numéricos se convierten en descripciones textuales estructuradas, asegurando que los datos sean representados de manera adecuada dentro del contexto de la generación de noticias.
2	Posteriormente, se lleva a cabo un formateo del texto para unificar la estructura de las entradas y garantizar una mayor coherencia. Este formateo incluye la eliminación de valores nulos o inconsistentes, así como la incorporación de etiquetas que permitan diferenciar la información estructurada de la noticia final generada.
3	Finalmente, la información procesada se prepara para su uso en el modelo mediante técnicas de tokenización, lo que permite dividir el texto en unidades más pequeñas, facilitando su interpretación por parte del modelo de lenguaje. La correcta aplicación de estos pasos es clave para optimizar el rendimiento y la precisión en la generación de textos automatizados.

Fuente: elaboración propia 2024.

5. Caracterización, optimización de parámetros y construcción de modelos de redacción de noticias

En esta sección se describirá el proceso de diseño e implementación y optimización de modelos de procesamiento de lenguaje natural (NLP) para la generación automática de noticias económicas. Esta etapa se centra en explorar las capacidades de los modelos GPT-2 y T5, destacando sus características técnicas, su capacidad para comprender y generar texto coherente y relevante, y su adecuación para el contexto de datos económicos.

5.1. Entrenamiento de los modelos

En este caso, no se entrena el modelo desde cero, sino que se emplea un enfoque de aprendizaje por transferencia utilizando modelos pre-entrenados (T5 o GPT-2). Esto permite aprovechar el conocimiento previo del modelo en la generación de lenguaje natural y adaptarlo a la tarea específica de generación de noticias económicas a partir de datos estructurados.

El proceso de entrenamiento se divide en las siguientes etapas:

1. Carga del Modelo Pre-entrenado

En lugar de inicializar un modelo desde cero, se utiliza un modelo previamente entrenado en grandes corpus de texto (como T5 o GPT-2). Estos modelos ya han aprendido una estructura lingüística general y pueden adaptarse más rápidamente a la tarea específica de generación de noticias económicas. Adicionalmente, se carga el modelo junto con su tokenizador correspondiente para asegurar que los datos de entrada sean procesados de manera adecuada.

2. Tokenización y Codificación de los Datos

Dado que los modelos de lenguaje requieren representaciones numéricas para procesar texto, se emplea un tokenizador que convierte las palabras en tokens y las transforma a identificadores numéricos.

El proceso se realiza de la siguiente manera:

- Se define una estructura textual a partir de los datos estructurados, combinando información como año, mes, indicador económico y variaciones.
- Esta estructura se convierte en texto natural mediante reglas predefinidas.
- Se utiliza el tokenizador del modelo para transformar el texto en una secuencia de tokens con longitud máxima definida.
- Se generan tensores de entrada que incluyen input_ids y attention_mask, los cuales indican qué partes del texto son relevantes para el modelo.

Para la salida esperada (las noticias económicas), se realiza un proceso similar, asegurando que las predicciones del modelo tengan una representación coherente con los ejemplos de entrenamiento.

3. Definición de los Hiperparámetros

Se establecen los parámetros clave que controlan el entrenamiento del modelo, incluyendo:

- Número de épocas: Define cuántas veces el modelo verá el conjunto de datos completo durante el entrenamiento.
- Tasa de aprendizaje: Determina qué tan rápido se ajustan los pesos del modelo durante la optimización.
- Tamaño del lote: Define cuántos ejemplos se procesan simultáneamente en cada actualización de parámetros.
- Técnicas de muestreo: Se configuran estrategias como beam search y top-p sampling para mejorar la coherencia y diversidad de las noticias generadas.

4. Ajuste de Pesos del Modelo

El modelo ajusta sus pesos utilizando un algoritmo de optimización basado en el error entre las predicciones generadas y los textos de referencia. El proceso sigue estos pasos:

- Cálculo de la pérdida: Se mide la discrepancia entre la salida del modelo y la noticia esperada mediante una función de pérdida, como la entropía cruzada.
- Retropropagación: Se calculan los gradientes de la función de pérdida con respecto a los pesos del modelo.
- Actualización de pesos: Se ajustan los pesos mediante un optimizador como Adam para minimizar la pérdida en las siguientes iteraciones.

Dado que el modelo no se entrena desde cero, este ajuste de pesos es incremental, refinando los conocimientos adquiridos previamente en el modelo base.

En conclusión, el uso de modelos pre-entrenados permite optimizar el proceso de generación de noticias sin necesidad de entrenar un modelo desde cero. Mediante el ajuste de pesos en un corpus especializado y la evaluación con métricas automáticas, se mejora la capacidad del modelo para generar texto relevante, preciso y contextualizado en el ámbito económico.

5.2. Modelo T5[3]

El modelo T5, basado en una arquitectura de codificador-decodificador, procesa datos económicos para generar texto estructurado. En el primer paso, el codificador toma la entrada, que en este caso es una secuencia textual construida a partir de datos económicos (por ejemplo, "Año: 2024, Mes: enero, País: Colombia, Indicador: Inflación, Valor_actual: 9%"), y la transforma en una representación interna de alta dimensionalidad. Este proceso utiliza múltiples capas de atención propia (self-attention), que permiten al modelo capturar las relaciones entre las palabras dentro de la entrada, como la asociación entre "Indicador" e "Inflación".

En el segundo paso, el decodificador genera una salida secuencial, utilizando la representación interna del codificador como entrada. A través de un mecanismo de atención cruzada (cross-attention), el decodificador consulta la información codificada para decidir qué tokens generar en cada paso. Por ejemplo, al procesar la representación de la entrada mencionada, el decodificador

[3] Los detalles del modelo implementado en Python pueden ser revisados en: https://github.com/LuViBeBe93/tesis/blob/main/T5_F1.ipynb

podría generar una oración como: "En enero de 2024, la inflación en Colombia alcanzó un 9% según el DANE, mostrando un crecimiento del 1% respecto al año anterior."

El entrenamiento del modelo se basa en la minimización de la pérdida de entropía cruzada, que mide la diferencia entre la secuencia de texto generada por el modelo y la secuencia de texto objetivo proporcionada en el conjunto de entrenamiento. Para optimizar esta pérdida, se utiliza el algoritmo Adam, un método de gradiente descendente adaptativo que ajusta dinámicamente las tasas de aprendizaje para cada parámetro del modelo. Por ejemplo, si el modelo genera "la inflación alcanzó un 8%" en lugar de "un 9%", la pérdida será alta, lo que inducirá actualizaciones en los pesos de las capas del modelo para mejorar su rendimiento.

En el proceso de generación de texto, el modelo T5 utiliza un enfoque secuencial y altamente optimizado. Una vez que el codificador ha transformado la entrada en una representación interna rica en contexto, el decodificador construye la salida o noticia económica paso a paso. Durante este proceso, el mecanismo de atención cruzada garantiza que cada token generado esté alineado con los datos originales. Por ejemplo, si los datos de entrada indican una inflación del 9%, el modelo asegura que esta información clave se incluya en la noticia generada.

El decodificador genera la salida de manera autoregresiva, eligiendo el token más probable en cada paso en función de los tokens generados previamente y la representación interna de la entrada. Este enfoque iterativo permite construir noticias coherentes y relevantes. Además, el uso de parámetros como temperatura y top-p aporta un equilibrio entre creatividad y precisión, asegurando que las narrativas sean naturales y contextualmente apropiadas.

Por ejemplo, con una entrada como: "Año: 2024, Mes: enero, Indicador: Inflación, Valor_actual: 9%, Variación: 1%, Entidad: DANE", el modelo podría generar la siguiente noticia: "En enero de 2024, la inflación en Colombia alcanzó un 9%, según reportó el DANE. Este incremento representa un aumento del 1% respecto al mismo mes del año anterior, reflejando una tendencia sostenida en el crecimiento de los precios al consumidor."

En resumen, el modelo T5 utiliza una combinación de mecanismos de atención, transformadores multicapa y técnicas de optimización avanzadas para generar textos coherentes y precisos a partir de datos estructurados, garantizando que las salidas reflejen fielmente las relaciones y tendencias en los datos económicos de entrada. Su capacidad para construir noticias relevantes, alineadas con los datos y estilísticamente naturales, lo convierte en una herramienta poderosa para la automatización de tareas de generación de texto.

5.3. Modelo GPT2[4]

El modelo GPT-2, basado en una arquitectura de transformadores unidireccionales, procesa datos económicos para generar texto estructurado y coherente. En el primer paso, los datos tabulares son preprocesados y convertidos en una representación textual. Por ejemplo, los datos: Año: 2024, Mes: enero, Indicador: Inflación, Valor_actual: 9%, Variación: 1%, Entidad: DANE, estos se transforman en la secuencia: "Año: 2024, Mes: enero, Indicador: Inflación, Valor_actual: 9%, Variación: 1%, Entidad: DANE Noticia: ...". Esta forma permite al modelo comprender la estructura y las relaciones inherentes en los datos.

En el segundo paso, el texto preprocesado se tokeniza, dividiéndolo en fragmentos (tokens) que representan palabras, símbolos o caracteres. Cada token es convertido en un índice numérico que corresponde a su posición en el vocabulario del modelo. Por ejemplo, el token "Inflación" podría ser representado como el índice 1453, y "9%" como 762. Esto convierte el texto de entrada en una secuencia de números lista para ser procesada por el modelo.

El tercer paso transforma estos tokens numéricos en vectores de alta dimensionalidad, llamados embeddings. Estos vectores encapsulan el significado y contexto de cada token, lo que permite al modelo trabajar con conceptos en lugar de simples números. Por ejemplo, el token "Inflación" podría tener una representación como [0.45, -0.12, 0.87, ...]. Estos vectores son refinados en múltiples capas de atención propia (self-attention), que evalúan las relaciones entre los tokens. En

[4] Los detalles del modelo implementado en Python pueden ser revisados en: https://github.com/LuViBeBe93/tesis/blob/main/GPT2.ipynb

este caso, el modelo podría determinar que "Indicador" está estrechamente relacionado con "Inflación" y "9%", pero menos relacionado con "Mes".

En el cuarto paso, el modelo utiliza los vectores procesados para generar texto. Cada token de salida es predicho uno a la vez, basado en los tokens anteriores y la representación interna del texto de entrada. Por ejemplo, si el modelo recibe la secuencia: "Año: 2024, Mes: enero, Indicador: Inflación, Valor_actual: 9%, Variación: 1%", esto podría generar la salida: "En enero de 2024, la inflación en Colombia alcanzó un 9%, reportada por el DANE, con un incremento del 1% respecto al año anterior."

Durante este proceso, el modelo GPT-2 utiliza su naturaleza autoregresiva, lo que significa que cada token generado se convierte en parte del contexto para predecir el siguiente. Este enfoque permite al modelo construir noticias de forma incremental y mantener la coherencia contextual en cada paso. Sin embargo, la generación también depende de los parámetros configurados, como la temperatura y el top-p, que controlan el grado de aleatoriedad en las salidas. Una temperatura más alta, por ejemplo, genera textos más creativos, pero con mayor riesgo de incoherencia, mientras que una temperatura baja produce resultados más conservadores y predecibles.

El entrenamiento del modelo se realiza minimizando la pérdida de entropía cruzada, que mide la discrepancia entre el texto generado y el texto objetivo en el conjunto de entrenamiento. Para optimizar esta pérdida, se utiliza el algoritmo Adam, un método adaptativo que ajusta las tasas de aprendizaje para mejorar la precisión del modelo.

En resumen, el modelo GPT-2 emplea un flujo de procesamiento que combina transformaciones textuales, tokenización, mecanismos de atención propia y optimización avanzada para generar textos coherentes y precisos a partir de datos estructurados. Este enfoque garantiza que las salidas reflejen fielmente las relaciones y tendencias en los datos económicos, facilitando la generación automática de reportes analíticos y narrativas informativas.

5.4. Optimización de parámetros

Los parámetros que usa de manera inicial o predeterminados el modelo T5 para la redacción de noticias son los siguientes:

Tabla 7. Listado de hiperparámetros iniciales del modelo T5

Hiperparámetros	Valor predeterinado
max_length: Define la longitud máxima de las secuencias de entrada y salida del modelo. De esta forma, trunca o rellena las secuencias para que todas tengan esta longitud.	128
batch_size: Cantidad de muestras procesadas en cada paso del entrenamiento.	2
num_epochs: Número de pasadas completas sobre el conjunto de datos de entrenamiento.	10
learning_rate: Tasa de aprendizaje, que controla cuánto se ajustan los pesos del modelo en cada paso de entrenamiento.	5e-5
num_beams: Número de trayectorias consideradas en el algoritmo de búsqueda para la generación de texto.	5
no_repeat_ngram_size: Evita la repetición de ngrams de tamaño especificado en el texto generado.	2
early_stopping: Detiene el proceso de generación si todos los haces han terminado antes de alcanzar la longitud máxima.	True
Temperature: Controla la aleatoriedad en la selección de palabras durante la generación de texto.	0.7
top_p: Establece un umbral acumulativo de probabilidad para la selección de palabras. En lugar de elegir entre todas las palabras posibles, selecciona solo aquellas cuya suma acumulativa de probabilidades está dentro del valor establecido.	0.9

Fuente: elaboración propia 2024.

Una vez conocidos los parámetros iniciales o predeterminados del modelo, se realizó una validación de las métricas de evaluación para poder tener insumos para una posterior comparación

con el mismo, pero con parámetros optimizados.

Tabla 8. Métricas versión inicial o básica del modelo T5

Métrica		***Resultado***
BLUE Score		*0.15*
BERTScore	*Precisión promedio*	*0.83*
	Recall promedio	*0.82*
	F1 promedio	*0.82*
Perplejidad		*1.09*

Fuente: elaboración propia 2024. Los resultados de estas métricas se encuentran en: https://github.com/LuViBeBe93/tesis/blob/main/T5_F2.ipynb

Por otra parte, los parámetros que usa de manera iniciales o predeterminados el modelo GPT2 para la redacción de noticias son los siguientes:

Tabla 9. Listado de hiperparámetros iniciales del modelo GPT2

Hiperparámetros	**Valor predeterminado**
max_length: Define la longitud máxima de las secuencias de entrada y salida del modelo. De esta forma, trunca o rellena las secuencias para que todas tengan esta longitud.	128
batch_size: Cantidad de muestras procesadas en cada paso del entrenamiento.	2
num_epochs: Número de pasadas completas sobre el conjunto de datos de entrenamiento.	10
num_beams: Número de trayectorias consideradas en el algoritmo de búsqueda para la generación de texto.	5
no_repeat_ngram_size: Evita la repetición de ngrams de tamaño especificado en el texto generado.	2
early_stopping: Detiene el proceso de generación si todos los haces han terminado antes de alcanzar la longitud máxima.	True

Temperature: Controla la aleatoriedad en la selección de palabras durante la generación de texto.	0.7
top_p: Establece un umbral acumulativo de probabilidad para la selección de palabras. En lugar de elegir entre todas las palabras posibles, selecciona solo aquellas cuya suma acumulativa de probabilidades está dentro del valor establecido.	0.9

Fuente: elaboración propia 2024.

Una vez conocidos los parámetros iniciales o predeterminados del modelo, se realizó una validación de las métricas de evaluación para poder tener insumos para una posterior comparación con el mismo, pero con parámetros optimizados.

Tabla 10. Métricas versión inicial o básica del modelo GPT2

Métrica		***Resultado***
BLUE Score		*0.126*
BERTScore	*Precisión promedio*	*0.59*
	Recall promedio	*0.83*
	F1 promedio	*0.69*
Perplejidad		*14.29*

Fuente: elaboración propia 2024. Los resultados de estas métricas se encuentran en: https://github.com/LuViBeBe93/tesis/blob/main/GPT2_F2.ipynb

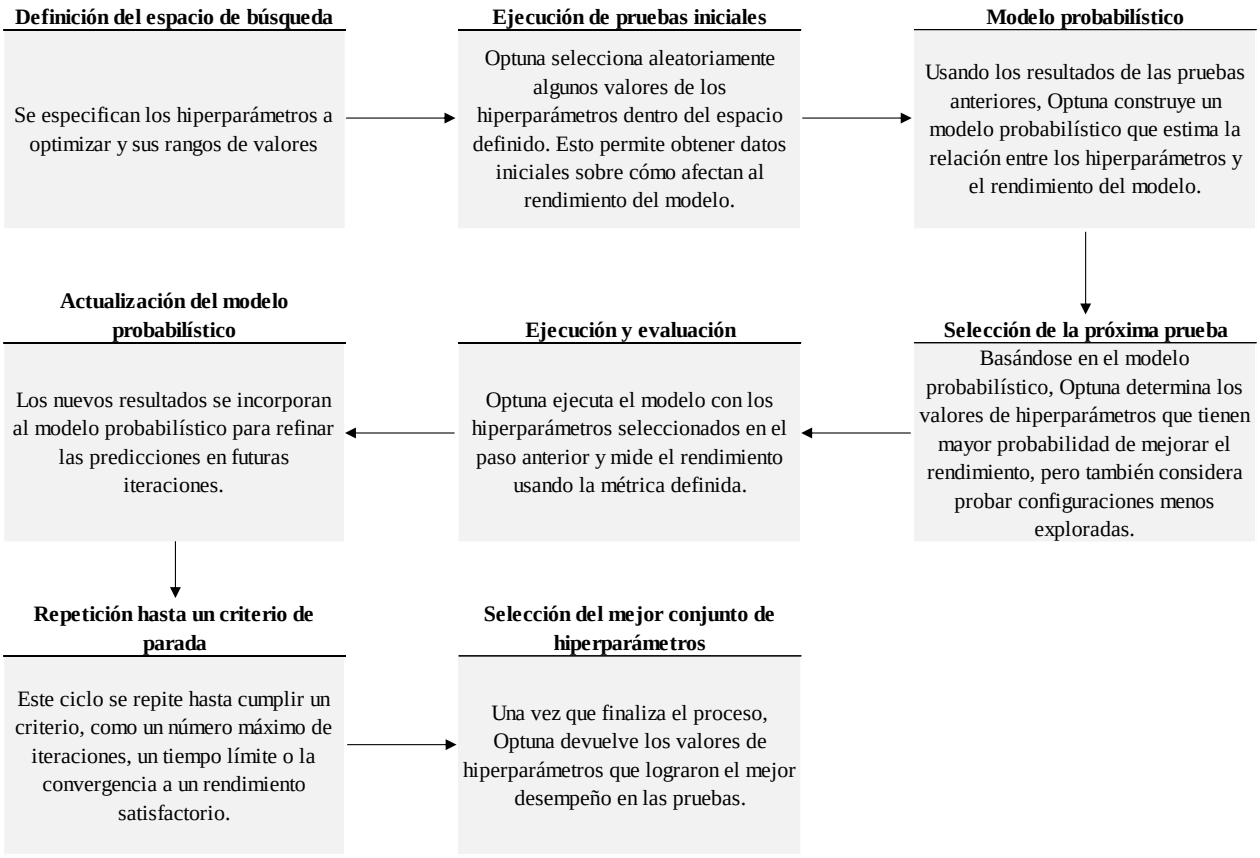
5.4.1. Optimización de parámetros modelo T5

En esta sección, se presenta el proceso de optimización de parámetros, el cual es un proceso clave en la configuración de los modelos predictivos y generativos, especialmente cuando se busca maximizar su desempeño en tareas específicas. En esta sección, se presentarán las metodologías de optimización por búsqueda bayesiana para los hiperpárametos y la selección por grilla para los parámetros de *temperature* y *top_p*. Estas técnicas buscan garantizar un balance óptimo entre precisión, velocidad y recursos computacionales.

Búsqueda bayesiana con Optuna

El proceso de búsqueda bayesiana con Optuna es una técnica que busca optimizar la selección de hiperparámetros de un modelo. Este proceso, funciona mediante un enfoque iterativo que equilibra la exploración de nuevas configuraciones y la explotación de las más prometedoras. El proceso que se maneja en la técnica se expone en la gráfica 2.

Gráfica 2. Proceso de búsqueda bayesiana con Optuna



Fuente: elaboración propia 2024.

Resultados sobre los hiperparámetros con Optuna

Una vez implementado el proceso de búsqueda bayesiana con Optuna sobre los hiperparámetros del modelo, se evaluaron 10 escenarios y se obtuvo un resultado máximo en la métrica BLEU en

la combinación que se resalta en color verde en la tabla 11, y estos fueron los valores óptimos para los hiperparámetros del modelo T5[5].

Tabla 11. Hiperparámetros evaluados con búsqueda bayesiana con Optuna

learning_rate	batch_size	num_epochs	num_beams	no_repeat_ngram_size	BLEU
2.90e-05	4	11	6	3	0.103090
1.18e-05	4	10	6	3	0.115323
0.00012	2	5	6	2	0.155628
4.53e-05	2	9	4	3	0.170505
1.41e-05	8	15	7	3	0.192376
7.83e-05	2	7	7	2	0.207575
4.84e-05	8	10	4	3	0.231622
5.28e-05	**8**	**14**	**4**	**3**	**0.241508**
3.47e-05	4	8	3	2	0.241508
0.00026	2	14	6	2	0.241508

Fuente: elaboración propia 2024. El código que contiene el procedimiento y resultados de la búsqueda bayesiana se encuentra en: https://github.com/LuViBeBe93/tesis/blob/main/T5_F3_V3.ipynb

Búsqueda por medio de grilla

La búsqueda por grilla es una técnica sistemática de optimización de parámetros en la que se exploran exhaustivamente todas las combinaciones posibles de valores dentro de un espacio predefinido. Este método construye una "grilla" de parámetros y evalúa el rendimiento del modelo en cada punto de dicha grilla. Una vez seleccionados los “hiperparámetros” en la sección anterior, se dispone a utilizar una grilla que construya las métricas descritas previamente para elegir la mejor combinación posible de *temperature* y *top_p* para el modelo. Los resultados son los siguientes:

Tabla 12. Parámetros de temperature y top_p evaluados con la grilla

Temperature	Top_p	bert_f1	bert_Precision	bert_Recall	Bleu	Perplejidad
0.5	0.5	0.80323279	0.80851358	0.80188811	0.21514718	1.1141828

[5] Este resultado se puede comparar con los valores iniciales que se presentaron en la tabla 4 y que hicieron parte de la primera ejecución del modelo.

0.5	0.6	0.83724988	0.83726603	0.84052855	0.30941666	1.1141828
0.5	0.7	0.83458096	0.83532196	0.8373484	0.30401243	1.1141828
0.5	0.8	0.83359212	0.83532047	0.8354544	0.29974201	1.1141828
0.5	0.9	0.83399242	0.83503371	0.83653408	0.30290334	1.1141828
0.7	0.5	0.83434808	0.83549398	0.83657604	0.30481222	1.1141828
0.7	0.6	0.83660924	0.83747369	0.83928448	0.31261021	1.1141828
0.7	0.7	0.83422929	0.83393389	0.83815688	0.30366479	1.1141828
0.7	0.8	0.83278412	0.83467937	0.83440655	0.29654689	1.1141828
0.7	0.9	0.83004713	0.832196	0.83142215	0.28737923	1.1141828
1	0.5	0.83141547	0.83202779	0.83447498	0.29326871	1.1141828
1	0.6	0.83429337	0.83607882	0.8359943	0.30056854	1.1141828
1	0.7	0.83544433	0.8378076	0.83644634	0.30144899	1.1141828
1	0.8	0.83031988	0.83305967	0.83108866	0.28558565	1.1141828
1	0.9	0.8230325	0.82648462	0.82323021	0.26267764	1.1141828
1.2	0.5	0.83403945	0.83455855	0.83708638	0.30008241	1.1141828
1.2	0.6	0.82733536	0.82996315	0.82840562	0.27953619	1.1141828
1.2	0.7	0.82750469	0.83137667	0.82718831	0.2777553	1.1141828
1.2	0.8	0.82696325	0.82942128	0.82789809	0.27201376	1.1141828
1.2	0.9	0.81267452	0.81739116	0.81206912	0.2309958	1.1141828
1.5	0.5	0.828107	0.8305636	0.8293615	0.28380957	1.1141828
1.5	0.6	0.82317787	0.82714468	0.82315969	0.26876856	1.1141828
1.5	0.7	0.81840163	0.82193202	0.81817847	0.24629197	1.1141828
1.5	0.8	0.80875409	0.81348634	0.80791521	0.21848278	1.1141828
1.5	0.9	0.79660237	0.79975682	0.79735601	0.19193426	1.1141828

Fuente: elaboración propia 2024. El código que contiene el procedimiento y resultados de la grilla se encuentra en: https://github.com/LuViBeBe93/tesis/blob/main/T5_F3_2.ipynb

Utilizando el BERT F1 como criterio de evaluación la mejor combinación de parámetros encontrada es la resaltada en color verde en la tabla 12.

5.4.2. Resultados con base en la optimización de parámetros T5

A través de un enfoque sistemático, se evalúan combinaciones de parámetros, midiendo su impacto en métricas de desempeño como BLEU, BERTScore y perplejidad. Este proceso permitió identificar la configuración óptima que equilibraba la creatividad en la generación de texto con la fidelidad a los datos originales, asegurando así resultados que cumplan con los objetivos del

proyecto.

Resultados[6]

Una vez aplicadas las métricas expuestas previamente, se obtuvieron los siguientes resultados para el modelo T5 con el que se redactaran las versiones finales de las noticias del capítulo posterior.

Tabla 13. Métricas con optimización de parámetros - modelo T5

Métrica		***Resultado***
BLUE Score		*0.14*
BERTScore	*Precisión promedio*	*0.84*
	Recall promedio	*0.82*
	F1 promedio	*0.83*
Perplejidad		*1.11*

Fuente: elaboración propia 2024.

En cuanto a los resultados obtenidos a través de las métricas evaluadas para el modelo T5, tanto en su versión inicial como después de la optimización de parámetros, se observa que, a pesar de una ligera disminución en el BLUE Score de 0.15 a 0.14 después de la optimización, la calidad general de las noticias generadas, en términos de coincidencia de n-gramas, permanece relativamente estable. El BLUE Score mide la coincidencia de n-gramas entre el texto generado y las referencias, por lo que una ligera disminución en este valor sugiere que la optimización no tuvo un impacto significativo en la alineación exacta de los términos del modelo con las fuentes de referencia. Esto podría implicar que, aunque los cambios en los parámetros del modelo optimizado no mejoraron sustancialmente las coincidencias exactas, el modelo pudo haber ganado en otros aspectos.

En cuanto a las métricas de BERTScore, se observa una ligera mejora en la precisión promedio de 0.83 a 0.84 y en el F1 promedio de 0.82 a 0.83. Estas mejoras reflejan que el modelo optimizado es más preciso en la identificación de elementos relevantes en la referencia (como términos clave) y presenta una ligera mejoría en la capacidad de equilibrar la precisión con la recuperación, lo que

[6] Los resultados de estas métricas se encuentran en https://github.com/LuViBeBe93/tesis/blob/main/T5_F5_op.ipynb

indica que las noticias generadas por el modelo son más relevantes y coherentes en relación con las referencias. Sin embargo, el Recall promedio se mantiene constante en 0.82, sugiriendo que la optimización no alteró la capacidad del modelo para recuperar la mayor parte de la información relevante.

El aumento en la perplejidad, de 1.09 a 1.11, sugiere que, aunque el modelo optimizado generó noticias con una mejor calidad semántica y de precisión, podría haber experimentado un pequeño aumento en la complejidad de los patrones generados. Este aumento en la perplejidad podría interpretarse como una ligera mayor dificultad en el modelo para predecir la siguiente palabra, lo cual no necesariamente es negativo, pero podría implicar una ligera mayor incertidumbre en la generación de texto.

5.4.3. Optimización de parámetros modelo GPT2

En esta sección, se presenta el proceso de optimización de parámetros, el cual es un proceso clave en la configuración de los modelos predictivos y generativos, especialmente cuando se busca maximizar su desempeño en tareas específicas. En esta sección, se presentarán las metodologías de optimización por búsqueda bayesiana para los hiperpárametos y la selección por grilla para los parámetros de *temperature* y *top_p*. Estas técnicas buscan garantizar un balance óptimo entre precisión, velocidad y recursos computacionales.

Búsqueda por medio de grilla

La búsqueda por grilla es una técnica sistemática de optimización de parámetros en la que se exploran exhaustivamente todas las combinaciones posibles de valores dentro de un espacio predefinido. Este método construye una "grilla" de parámetros y evalúa el rendimiento del modelo en cada punto de dicha grilla. Aquí se construyen las métricas descritas previamente para elegir la mejor combinación posible de *parámetros* para el modelo. Los resultados son los siguientes:

Resultados sobre los hiperparámetros con Optuna

Una vez implementado el proceso de búsqueda bayesiana con Optuna sobre los hiperparámetros del modelo, se evaluaron 10 escenarios y se obtuvo un resultado máximo en la métrica BLEU en la combinación que se resalta en color verde en la tabla 14, y estos fueron los valores óptimos para los hiperparámetros del modelo GPT2[7].

Tabla 14. Hiperparámetros evaluados con búsqueda bayesiana con Optuna

learning_rate	batch_size	num_epochs	num_beams	no_repeat_ngram_size	BLEU
2.97e-05	**4**	**7**	**3**	**3**	**0.554725**
1.44e-05	2	8	5	3	0.554725
1.70e-05	2	5	5	2	0.554725
1.39e-05	4	6	5	3	0.554725
9.94e-05	8	9	7	2	0.554725
0.00023	8	15	3	2	0.554725
0.00037	4	9	4	2	0.554725
0.00012	8	13	7	2	0.554725
0.00017	2	7	5	3	0.554725
4.10e-05	4	15	7	2	0.554725

Fuente: elaboración propia 2024. El código que contiene el procedimiento y resultados de la búsqueda bayesiana se encuentra en: https://github.com/LuViBeBe93/tesis/blob/main/GPT2_F3_V3.ipynb

Búsqueda por medio de grilla

La búsqueda por grilla es una técnica sistemática de optimización de parámetros en la que se exploran exhaustivamente todas las combinaciones posibles de valores dentro de un espacio predefinido. Este método construye una "grilla" de parámetros y evalúa el rendimiento del modelo en cada punto de dicha grilla. Una vez seleccionados los “hiperparámetros” en la sección anterior, se dispone a utilizar una grilla que construya las métricas descritas previamente para elegir la mejor combinación posible de *temperature* y *top_p* para el modelo. Los resultados son los siguientes:

[7] Este resultado se puede comparar con los valores iniciales que se presentaron en la tabla 10 y que hicieron parte de la primera ejecución del modelo.

Tabla 15. Parámetros de temperature y top_p evaluados con la grilla

Temperature	Top_p	bert_f1	bert_Precision	bert_Recall	Bleu	Perplejidad
0.5	0.5	0.64606	0.550258	0.785812	0.14365	1.212096
0.5	0.6	0.646439	0.550633	0.786174	0.14444	1.212096
0.5	0.7	0.646498	0.550544	0.78656	0.143404	1.212096
0.5	0.8	0.646576	0.55087	0.786119	0.143913	1.212096
0.5	0.9	0.646549	0.550571	0.786749	0.145101	1.212096
0.7	0.5	0.646494	0.550455	0.786762	0.143719	1.212096
0.7	0.6	0.646288	0.550505	0.785994	0.143606	1.212096
0.7	0.7	0.646454	0.550486	0.786586	0.143029	1.212096
0.7	0.8	0.646617	0.550699	0.786658	0.145313	1.212096
0.7	**0.9**	**0.646899**	**0.551084**	**0.786642**	**0.146405**	**1.212096**
1	0.5	0.646296	0.550458	0.786112	0.144454	1.212096
1	0.6	0.646545	0.550651	0.786488	0.144441	1.212096
1	0.7	0.646596	0.55059	0.786859	0.144359	1.212096
1	0.8	0.646738	0.550655	0.787078	0.14648	1.212096
1	0.9	0.646086	0.550304	0.785681	0.145935	1.212096
1.2	0.5	0.646514	0.550532	0.786689	0.143815	1.212096
1.2	0.6	0.646508	0.550633	0.786458	0.145035	1.212096
1.2	0.7	0.646743	0.550885	0.786578	0.145082	1.212096
1.2	0.8	0.646005	0.55044	0.7853	0.145446	1.212096
1.2	0.9	0.645097	0.54973	0.7839	0.142576	1.212096
1.5	0.5	0.646609	0.550645	0.78668	0.144391	1.212096
1.5	0.6	0.646671	0.550835	0.786523	0.145549	1.212096
1.5	0.7	0.646621	0.550793	0.786392	0.147084	1.212096
1.5	0.8	0.64526	0.549882	0.784194	0.14236	1.212096
1.5	0.9	0.645201	0.549525	0.784725	0.141275	1.212096

Fuente: elaboración propia 2024. El código que contiene el procedimiento y resultados de la grilla se encuentra en: https://github.com/LuViBeBe93/tesis/blob/main/GPT2_F3_V2.ipynb

Utilizando el BERT F1 como criterio de evaluación la mejor combinación de parámetros encontrada es la resaltada en color verde en la tabla 15.

5.4.4. Resultados con base en la optimización de parámetros GPT2

En esta sección, se busca analizar y ajustar los parámetros clave del modelo para maximizar su

rendimiento en términos de calidad, coherencia y diversidad. A través de un enfoque sistemático, se evalúan combinaciones de parámetros, midiendo su impacto en métricas de desempeño como BLEU, BERTScore y perplejidad. Este proceso permite identificar la configuración óptima que equilibre la creatividad en la generación de texto con la fidelidad a los datos originales, asegurando así resultados que cumplan con los objetivos del proyecto.

Resultados[8]

Una vez aplicadas las métricas expuestas previamente, se obtuvieron los siguientes resultados para el modelo GPT2 con el que se redactaran las versiones finales de las noticias del capítulo posterior.

Tabla 16. Métricas con optimización de parámetros - modelo GPT2

Métrica		***Resultado***
BLUE Score		*0.152*
BERTScore	*Precisión promedio*	*0.60*
	Recall promedio	*0.84*
	F1 promedio	*0.70*
Perplejidad		*12.75*

Fuente: elaboración propia 2024.

En comparación con los resultados obtenidos con el modelo T5, en el modelo GPT-2, los resultados de las métricas presentan un panorama diferente. El BLUE Score aumenta de 0.126 a 0.152 después de la optimización, lo que indica una mejora en la generación de texto coherente y alineado con las referencias. Esta mejora en el BLUE Score sugiere que el modelo optimizado logra generar noticias que se alinean mejor con las expectativas de coincidencia de n-gramas, reflejando una mayor capacidad para replicar estructuras similares a las de las fuentes de referencia.

En cuanto a las métricas de BERTScore, la mejora en la precisión promedio de 0.59 a 0.60, aunque modesta, implica que el modelo optimizado logra una leve mejora en la identificación de elementos

[8] Los resultados de estas métricas se encuentran en: https://github.com/LuViBeBe93/tesis/blob/main/GPT2_F5_op.ipynb

clave dentro del contexto de las noticias generadas. Por su parte, el recall promedio mejora de 0.83 a 0.84, lo que refleja un pequeño aumento en la capacidad del modelo para recuperar términos importantes dentro del texto, lo que mejora la relevancia de la salida generada en relación con el contenido esperado.

La mejora más destacada en el modelo GPT-2 se observa en la perplejidad, que disminuye significativamente de 14.29 a 12.75. Este cambio es indicativo de una reducción en la complejidad de la generación de texto, lo que sugiere que el modelo optimizado tiene una mayor capacidad para predecir correctamente las palabras que siguen en el contexto, produciendo salidas más fluidas y coherentes. Esta mejora en la perplejidad implica que el modelo optimizado genera noticias económicas con mayor coherencia interna y menor incertidumbre en la predicción de las palabras.

5.5. Evaluación del desempeño en la generación de noticias

En la evaluación del modelo de generación de noticias económicas, hemos implementado un conjunto de métricas que capturan aspectos clave del desempeño:

Fluidez y coherencia

Utilizamos Perplejidad, una métrica estándar para modelos de lenguaje, que permite medir qué tan bien el modelo predice secuencias de palabras plausibles en español. Una perplejidad baja sugiere que el modelo genera texto estructurado y gramaticalmente correcto.

Similitud semántica con textos de referencia

Aplicamos BERTScore, que, a diferencia de métricas tradicionales basadas en coincidencias exactas de palabras, evalúa la similitud semántica entre la noticia generada y la referencia. Esto nos ayuda a medir qué tan bien el modelo captura la esencia del mensaje esperado, más allá de la coincidencia exacta de términos.

Fidelidad a los datos de entrada

La forma en que estructuramos los datos de entrada y los textos generados permite evaluar de manera indirecta la coherencia factual del modelo. Si bien no tenemos una métrica específica para esto, el diseño de entrada y salida nos da una base para verificar si los valores y tendencias

económicas están siendo correctamente interpretados por el modelo.

Medición de similitud léxica y estructura del texto

El BLEU Score es útil para evaluar el grado de coincidencia entre la noticia generada y el texto de referencia, lo que ayuda a medir qué tan bien el modelo sigue patrones conocidos en la redacción de noticias económicas.

Por otra parte, si bien estas métricas ya nos proporcionan un análisis sólido sobre la calidad del modelo, en futuras investigaciones podríamos complementar la evaluación con otros enfoques que profundicen en el desempeño de la generación de noticias:

Verificación factual automática

Incorporar herramientas de detección de errores numéricos y de coherencia con los datos de entrada, como FactCC o modelos específicos de verificación factual en español.

Medición de diversidad y originalidad

Aplicar métricas como Dist-1, Dist-2 para evaluar la diversidad en el texto generado y evitar estructuras repetitivas en la redacción. Usar Self-BLEU para medir qué tan variado es el contenido generado en comparación con otras noticias dentro del mismo conjunto de datos.

Evaluación de estilo periodístico

Desarrollar modelos supervisados que analicen si el texto generado cumple con normas de redacción periodística en economía, considerando tono, estructura y claridad.

En conclusión, el conjunto de métricas que ya aplicamos nos permite evaluar aspectos fundamentales del desempeño del modelo, incluyendo fluidez, similitud semántica y coherencia textual. Sin embargo, futuras investigaciones podrían explorar métricas adicionales para evaluar la precisión factual y la diversidad del contenido generado. Esto permitiría avanzar hacia una evaluación aún más integral del desempeño en la generación automática de noticias económicas.

6. Redacción de textos, criterios de comparación y evaluación de los modelos implementados

6.1. Redacción modelo T5 sin optimizar

De acuerdo con el esquema planteado en la sección 4.1.1 de entrada de información y salida de una noticia económica, el modelo T5, basado en su capacidad de generar texto a partir de datos estructurados, fue utilizado para **redactar tres ejemplos de noticias económicas por cada conjunto de datos de entrada** (ver tabla 15), es importante mencionar que estos ejemplos fueron construidos a partir de la forma predeterminada del modelo y sin optimizar parámetros. Por otra parte, el modelo emplea un enfoque probabilístico en la generación de texto, permitiendo variabilidad en las salidas mediante el uso de parámetros como *temperature* y *top-p*. Estos parámetros introducen un grado de aleatoriedad controlada, priorizando tokens más probables, pero explorando alternativas para enriquecer la creatividad y diversidad en los resultados.

Cada ejemplo generado refleja una narrativa diferente, adaptada a los datos económicos de entrada. Por ejemplo, para un conjunto de datos que describe un aumento en la inflación mensual en Colombia, el primer texto podría resaltar el crecimiento porcentual en comparación con el año anterior, el segundo enfatizar tan solo en la cifra del periodo de análisis, y el tercero incluir tan solo la variación y alguna consecuencia de la misma. Esta diversidad en los ejemplos garantiza que las noticias sean útiles para distintos propósitos comunicativos, como análisis técnico, divulgación general o contenido mediático.

Tabla 17. Noticias redactadas con el modelo T5

Información de entrada	Noticia generada
Año: 2024 Mes: enero Frecuencia: Mensual Año_comparación: 2023	*Noticia generada:* El DANE informó que la inflación en Colombia fue del 9%.

País: Colombia Tendencia: Crecimiento Sector: Economía Indicador: **Inflación** Valor_actual: 9% Variación: 1% Valor_comparación: 8% Entidad: DANE	*Noticia generada:* El DANE reportó que la inflación anual en enero de 2024 fue del 9%. *Noticia generada:* En enero de 2024, la tasa de inflación alcanzó el 9%, lo que representó un incremento respecto al año 2023.
Año: 2024 Trimestre: Primer Frecuencia: Trimestral País: Argentina Tendencia: Crecimiento Sector: Economía **Indicador: PIB** Valor_actual: 10%	*Noticia generada:* El Producto Interno Bruto (PIB) de Argentina crece un 10% en el primer trimestre de 2024 *Noticia generada:* Crecimiento del PIB de Argentina: 10% en el primer trimestre de 2024 *Noticia generada:* El PIB de Argentina creció un 10% en el primer trimestre de 2024, superando las expectativas del mercado.
Año: 2020 Mes: mayo Frecuencia: Mensual País: Bolivia Tendencia: Disminuye Sector: Economía **Indicador: Tasa_desempleo** Valor_actual: 3%	*Noticia generada:* Para mayo de 2020, la tasa de desempleo en Bolivia fue del 3%. *Noticia generada:* La tasa de desempleo en Bolivia bajó al 3%, lo que representó un descenso frente al año anterior. *Noticia generada:* Para mayo de 2020, la tasa de desempleo en Bolivia descendió a 3%.
Año: 1993 Frecuencia: Anual Año_comparación: 1992 País: Nicaragua Tendencia: Disminuye Sector: Economía **Indicador: Importaciones** Valor_actual: 12000 Variación: 1%	*Noticia generada:* En 1993, las importaciones en Nicaragua sumaron 12000 millones de dólares, lo que representó una disminución del 1% respecto al año 1992 *Noticia generada:* El año de 1993, las importaciones en Nicaragua sumaron 12000 millones de dólares, lo que representó una disminución del 1%. *Noticia generada:* En 1993 las importaciones en Nicaragua sumaron 12000 millones de dólares, mostrando una caída del 1% respecto al año 1992

Año: 1993 Frecuencia: Anual Año_comparación: 1992 País: Uruguay Tendencia: Disminuye Sector: Economía **Indicador: Exportaciones** Valor_actual: 10000 Variación: 1%	*Noticia generada:* Las exportaciones de Uruguay alcanzan los 10000 millones de dólares en 1993, una disminución del 1% respecto al año 1992 *Noticia generada:* Exportaciones alcanzan los 10000 millones de dólares en 1993, una caída del 1% respecto al año de 1992 *Noticia generada:* Exportaciones de Uruguay alcanzó los 10000 millones de dólares en 1993
Año: 2015 Mes: noviembre Frecuencia: Mensual País: Chile Tendencia: Disminuye Sector: Economía **Indicador: Precio_petróleo** Valor_actual: 81	*Noticia generada:* El precio del barril de petróleo alcanzó los 81 dólares en noviembre de 2015 *Noticia generada:* El precio del petróleo alcanzó unos 81 dólares por barril en noviembre de 2015 *Noticia generada:* El precio del petróleo alcanzó los 81 dólares en noviembre de 2015
Año: 2015 Mes: noviembre Frecuencia: Mensual País: Italia Tendencia: Disminuye Sector: Economía **Indicador: Inversión_extranjera_directa** Valor_actual: 19000	*Noticia generada:* La IED en Italia alcanzó los 19000 millones de dólares *Noticia generada:* La inversión extranjera directa en Italia alcanzó los 19000 millones de dólares en noviembre de 2015, una disminución del mercado. *Noticia generada:* La inversión extranjera directa (IED) en Italia alcanzó el 19000 millones de dólares a noviembre de 2015

Fuente: elaboración propia 2024. El código que contiene los ejemplos antes descritos se encuentra en: https://github.com/LuViBeBe93/tesis/blob/main/T5_F1.ipynb

6.2. Redacción modelo GPT2 sin optimizar

En la tabla 18, se presentan ejemplos de noticias generadas con el modelo GPT-2, el cual fue utilizado en su forma predeterminada y sin optimizar sus parámetros. En cuanto a los ejemplos, estos ilustran cómo, a partir de los mismos datos de entrada, se obtienen redacciones distintas debido a la naturaleza probabilística del modelo y los parámetros configurados para la generación de texto. Cada noticia tiene ligeras variaciones en la presentación de la información y los detalles incluidos. Estas diferencias surgen debido a factores como:

Parámetros de generación: La configuración de valores como la temperatura y el top-p influye en la creatividad del modelo. Una temperatura más alta permite salidas más variadas, mientras que una temperatura baja produce textos más conservadores y repetitivos.
Complejidad del modelo: Durante la generación, el modelo selecciona palabras basándose en probabilidades. Esto puede resultar en enfoques diversos sobre qué aspectos enfatizar (por ejemplo, variaciones porcentuales o factores que impactan la inflación).
Contexto y truncamiento: Algunas generaciones pueden omitir información o incluir detalles irrelevantes debido al límite de tokens o al manejo de la atención.

En los ejemplos dados, se observa cómo el modelo ajusta ligeramente los porcentajes, incluye explicaciones diferentes sobre el descenso de la inflación, e incluso cambia la estructura de la noticia. Esto subraya la capacidad del GPT-2 para explorar múltiples redacciones posibles mientras mantiene la coherencia en el contexto económico proporcionado.

Tabla 18. Noticias redactadas con el modelo GPT2

Información de entrada	Noticia generada
Año: 2024 Mes: enero Frecuencia: Mensual Año_comparación: 2023 País: Colombia Tendencia: Crecimiento	*Noticia generada:* La tasa de inflación en Colombia se moderó al 9.2% en enero de 2024. Este descenso se debe principalmente a la caída en los precios de alimentos y energía. se beneficiarán principalmente con el aumento de las

Sector: Economía Indicador: **Inflación** Valor_actual: 9% Variación: 1% Valor_comparación: 8% Entidad: DANE	*Noticia generada:* La tasa de inflación en Colombia se moderó al 9% en enero de 2024. Este descenso fue impulsado principalmente por la caída en los precios de alimentos y energía. 8% de las ventas externas de Colombia fueron de 8 *Noticia generada:* La tasa de inflación en Colombia se situó en 9.1% en enero de 2024, una desaceleración en comparación con el 8.7% de 20.4% registrado en el mismo mes del año 2023.
Año: 2024 Trimestre: Primer Frecuencia: Trimestral País: Argentina Tendencia: Crecimiento Sector: Economía **Indicador: PIB** Valor_actual: 10%	*Noticia generada:* La economía de Argentina creció a un ritmo anualizado del 10.3% en el primer trimestre de 2023, superando las expectativas del mercado. Este crecimiento fue impulsado principalmente por el crecimiento de la actividad en los sectores de minería, comercio y servicios. *Noticia generada:* El Producto Interno Bruto (PIB) de Argentina creció un 10% anual en el primer trimestre de 2023, superando las expectativas del mercado., impulsada por el crecimiento de la actividad en los sectores de comercio exterior y el incremento de las exportaciones por un *Noticia generada:* La economía de Argentina creció a un ritmo anualizado del 10% en el primer trimestre de 2023, superando las expectativas del mercado. Este crecimiento fue impulsado principalmente por el crecimiento de la actividad en los sectores de comercio y la recuperación del consumo interno.
Año: 2020 Mes: mayo Frecuencia: Mensual País: Bolivia Tendencia: Disminuye Sector: Economía **Indicador: Tasa_desempleo** Valor_actual: 3%	*Noticia generada:* La tasa de desempleo en Bolivia descendió al 3.0% en mayo de 2020 tuvo una disminución de 0.8% puntos por encima de la tasa registrada en el mismo mes del año anterior, reflejando una tendencia a la baja del comercio exterior del *Noticia generada:* La tasa de desempleo en Bolivia baja al 3% en mayo de 2020 se ubica en 3.2% lo que indica una desaceleración en comparación con el mismo mes del año anterior. cae al 10.3% registrado en

	Noticia generada: La tasa de desempleo en Bolivia se disminuyó al 3.0% en mayo de 2020 tuvo una disminución de 0.8% puntos por encima de la tasa registrada en el mismo mes del año anterior. Este descenso se debe principalmente a la caída en
Año: 1993 Frecuencia: Anual Año_comparación: 1992 País: Nicaragua Tendencia: Disminuye Sector: Economía **Indicador: Importaciones** Valor_actual: 12000 Variación: 1%	*Noticia generada:* Las importaciones en Nicaragua cayeron un 1.3% en 1993 respecto al mismo mes del año 1992, reflejando una tendencia a la baja del comercio exterior del país. se registró una reducción en los precios en comparación con el del mismo período de 1992 IPC *Noticia generada:* Las importaciones en Nicaragua cayeron un 10% en 1993 respecto al mismo año 1992izaron 12.0 millones de dólares nan, se presentan en el informe sobre el comercio exterior de Nicaragua se proyecta que se ubique en 13.1, se espera que el *Noticia generada:* Nicaragua experimentó una disminución de las importaciones del 10.1% en 1993 respecto al año 1992 respecto a el mismo mes del año anterior. nan, se ubique en 11174.0, de acuerdo con la información de exportaciones del gobierno de Nicaragua.
Año: 1993 Frecuencia: Anual Año_comparación: 1992 País: Uruguay Tendencia: Disminuye Sector: Economía **Indicador: Exportaciones** Valor_actual: 10000 Variación: 1%	*Noticia generada:* En 1993 las exportaciones en Uruguay sumaron 10000.0, registando así una caída del 0.1% respecto al año 1992 tuvo una disminución de 2.5%. 6% en comparación con el mismo mes de 1992 se ubicó en los 100 *Noticia generada:* Las exportaciones en Uruguay cayeron un 10% en 1993, un descenso en comparación con el 20% registrado en el mismo mes de 1992 Boliviana se presentó una reducción en la producción de crudo por parte de los países de la OPEP+ y las *Noticia generada:* Las exportaciones en Uruguay cayeron un 10% en 1993, reflejando una tendencia a la baja en el comercio exterior del país. se ubica en 10.6% sobre el nivel de ingresos de las exportaciones de baja por parte de los países de la OP
Año: 2015 Mes: noviembre	*Noticia generada:* El precio del petróleo en Chile muestra una tendencia hacia la baja, en parte debido a factores como la menor

Frecuencia: Mensual País: Chile Tendencia: Disminuye Sector: Economía **Indicador: Precio_petróleo** Valor_actual: 81	demanda global y la transición hacia energías renovables, que afecta el mercado de combustibles fósiles y el comercio exterior de Chile. Las tensiones geopolíticas en Ucrania *Noticia generada:* En noviembre de 2015 se espera que el precio del petróleo en Chile llegue a los 81 dólares por barril, un aumento del 11% en comparación con el mes anterior., la cual se presentó en el informe de balanza comercial, mostró una reducción al *Noticia generada:* En noviembre de 2015 el precio del petróleo en Chile muestra una tendencia hacia la baja, en parte debido a factores como la menor demanda global y la transición hacia energías renovables, que afecta el mercado de combustibles fósiles. Este es un descenso en comparación con el 11
Año: 2015 Mes: noviembre Frecuencia: Mensual País: Italia Tendencia: Disminuye Sector: Economía **Indicador: Inversión_extranjera_directa** Valor_actual: 19000	*Noticia generada:* La IED en Italia alcanzó los 190000 millones de dólares en noviembre de 2015 con una variación de 0.4% por debajo de la tasa registrada en el mismo mes del año 2023. Este es un descenso en comparación con *Noticia generada:* La inversión extranjera directa (IED) en Italia alcanzó los 190000 millones de dólares en noviembre de 2015 lo que representa un crecimiento del 10% en comparación con el mismo mes del año anterior. reflejando una tendencia a la *Noticia generada:* La inversión extranjera directa (IED) en Italia disminuyó un 20% en noviembre de 2015 hasta los 19200 millones de dólares. se presentó una reducción en el nivel de ingresos de inversión privada se recuperará en octubre de

Fuente: elaboración propia 2024. El código que contiene los ejemplos antes descritos se encuentra en: https://github.com/LuViBeBe93/tesis/blob/main/GPT2.ipynb

6.3. Redacción final modelos T5 y GPT2 con parámetros optimizados

En esta sección se presentan los resultados obtenidos tras la optimización de parámetros en los modelos T5 y GPT-2 para la redacción de noticias económicas. La optimización se enfocó en ajustar los parámetros de cada modelo con el objetivo de mejorar su coherencia, precisión y relevancia contextual. A continuación, se presentan los textos redactados por los modelos una vez incluidos los parámetros obtenidos en el proceso de optimización.

Tabla 19. Noticias redactadas con los modelos T5 y GPT2 con parámetros optimizados

Información de entrada	Noticia generada modelo T5	Noticia generada modelo GPT2
Año: 2024 Mes: enero Frecuencia: Mensual Año_comparación: 2023 País: Colombia Tendencia: Crecimiento Sector: Economía Indicador: **Inflación** Valor_actual: 9% Variación: 1% Valor_comparación: 8% Entidad: DANE	*Noticia generada:* El DANE, presentó hoy el informe sobre la inflación en enero de 2024, correspondiente a un crecimiento de 1 % en comparación con los 8% del año anterior.	*Noticia generada:* En Colombia, el Dane informó que en enero de 2024 la variación mensual de la inflación fue de 9.0% en su medición anual..0 millones de dólares de inversión extranjera directa (IED) en Colombia
Año: 2024 Trimestre: Primer Frecuencia: Trimestral País: Argentina Tendencia: Crecimiento Sector: Economía **Indicador: PIB** Valor_actual: 10%	*Noticia generada:* El PIB de Argentina creció un 10% en el primer trimestre de 2024.	*Noticia generada:* La economía de Argentina creció un 10% en el primer trimestre de 2024 en comparación con el mismo período del año anterior., Banco Mundial _ millones de dólares en un primer trimestre, un crecimiento del 10% respecto al mismo

Año: 2020 Mes: mayo Frecuencia: Mensual País: Bolivia Tendencia: Disminuye Sector: Economía **Indicador: Tasa_desempleo** Valor_actual: 3%	*Noticia generada:* La tasa de desempleo en Bolivia disminuye a 3% en mayo de 2020	*Noticia generada:* La tasa de desempleo en Bolivia descendió al 3% en mayo de 2020 sempleo: 0.096, Variaciones: nan, Valor: nan Farnes en su medición anual se ubicó en 9.6
Año: 1993 Frecuencia: Anual Año_comparación: 1992 País: Nicaragua Tendencia: Disminuye Sector: Economía **Indicador: Importaciones** Valor_actual: 12000 Variación: 1%	*Noticia generada:* Las importaciones en Nicaragua llegaron a los 12000 millones de dólares en 1993, un crecimiento del 1% en comparación con el mes de 1992	*Noticia generada:* Las importaciones en Nicaragua cayeron un 1% en 1993 en comparación con el mismo periodo de 1992 se ubicaron en los 12000 millones de dólares se ubicó en los sectores de Economía y comercio exterior del país.
Año: 1993 Frecuencia: Anual Año_comparación: 1992 País: Uruguay Tendencia: Disminuye Sector: Economía **Indicador: Exportaciones** Valor_actual: 10000 Variación: 1%	*Noticia generada:* Las exportaciones de Uruguay llegaron a 10000 millones de dólares en 1993, un crecimiento del 1% respecto al mismo periodo de 1992	*Noticia generada:* Las exportaciones en Uruguay cayeron un 10% en 1993, alcanzando un valor de 10000 dólares por barril. en Uruguay se recuperará en 1993 con un aumento proyectado del 11.5%.
Año: 2015 Mes: noviembre Frecuencia: Mensual País: Chile Tendencia: Disminuye Sector: Economía **Indicador: Precio_petróleo** Valor_actual: 81	*Noticia generada:* El precio del petróleo en Chile se moderó a los 81 dólares por barril en noviembre de 2015	*Noticia generada:* En noviembre de 2015, el precio del petróleo en Chile muestra una tendencia hacia la baja, en parte debido a factores como la menor demanda global y la transición hacia energías renovables, que afecta el mercado de combustibles fósiles en Chile durante el mes.

Año: 2015 Mes: noviembre Frecuencia: Mensual País: Italia Tendencia: Disminuye Sector: Economía **Indicador: Inversión_extranjera_directa** Valor_actual: 19000	*Noticia generada:* La inversión extranjera directa (IED) en Italia alcanzó los 19000 millones de dólares en noviembre de 2015	*Noticia generada:* La IED en Italia alcanzó los 19000 millones de dólares en noviembre de 2015, lo que representa un crecimiento del 10.7% en comparación con el mismo mes del año anterior.

Fuente: elaboración propia 2024.

El código que contiene los ejemplos antes descritos para el modelo T5 se encuentra en: https://github.com/LuViBeBe93/tesis/blob/main/T5_F4.ipynb. El código que contiene los ejemplos antes descritos para el modelo GPT2 se encuentra en: https://github.com/LuViBeBe93/tesis/blob/main/GPT2_F4.ipynb

Los resultados obtenidos en la comparación entre los textos generados por los modelos T5 y GPT-2 demuestran diferencias significativas en cuanto a precisión, coherencia y relevancia contextual, lo que resalta las ventajas y limitaciones de cada modelo en la tarea de generación de noticias económicas.

En primer lugar, el modelo T5 superó a GPT-2 en métricas clave como BERTScore, BLEU y perplejidad. Esto se debe principalmente a que T5 utiliza una arquitectura de codificador-decodificador, lo que le permite procesar y entender mejor las relaciones entre las palabras de entrada y generar una salida que es más coherente y semánticamente similar al texto objetivo. Por ejemplo, en los casos donde la entrada describe cambios en la inflación, T5 generó textos que no solo mencionaban el porcentaje de variación, sino que también incluían información contextual adicional, como el año de comparación o el impacto de la variación en la economía.

Por otro lado, GPT-2, basado en un enfoque unidireccional, mostró una mayor propensión a desviarse del contexto inicial, lo que puede explicarse por la naturaleza probabilística de su generación de texto. Este comportamiento es evidente en algunos ejemplos donde el modelo incluyó información incoherente o detalles irrelevantes que no estaban presentes en los datos de entrada. A pesar de esto, GPT-2 destacó en la generación de textos más creativos y variados, lo que

podría ser útil en aplicaciones donde la creatividad tenga mayor peso sobre la fidelidad al contexto.

Además, la diferencia en la calidad de las salidas también puede atribuirse a los mecanismos de atención utilizados por cada modelo. T5 emplea atención cruzada (cross-attention), lo que le permite enfocar directamente los datos de entrada al momento de generar texto. En contraste, GPT-2 utiliza atención propia (self-attention) únicamente sobre las palabras generadas previamente, lo que reduce su capacidad para anclar las salidas al contexto inicial.

Finalmente, estos resultados destacan la importancia de la selección del modelo según los objetivos de la tarea. T5 es más adecuado para contextos donde la precisión y la fidelidad al dato son esenciales, como informes económicos o análisis técnicos. Por su parte, GPT-2 podría ser más útil en escenarios que requieran creatividad, como la redacción de contenido promocional o narrativas de carácter general.

En resumen, ambos modelos tienen aplicaciones valiosas dependiendo del enfoque deseado, pero la elección de uno sobre el otro debe basarse en las necesidades específicas de coherencia, precisión y creatividad del proyecto.

7. Conclusiones

En primer lugar, el proyecto ha cumplido con su objetivo principal: *generar noticias a partir de conjuntos de datos económicos, utilizando técnicas de inteligencia artificial y aprendizaje automático*. Mediante el uso de los modelos T5 y GPT-2, se logró transformar datos económicos complejos en narrativas claras, estructuradas y relevantes. Ambos modelos fueron ajustados y evaluados en términos de métricas clave, garantizando que los textos generados reflejen fielmente las tendencias y relaciones presentes en los datos de entrada. Este logro representa un avance significativo en la automatización de la generación de contenido periodístico, mostrando cómo las técnicas de procesamiento de lenguaje natural pueden ser aplicadas eficazmente en el ámbito económico.

El modelo T5 demostró ser una herramienta altamente eficaz para la generación de textos económicos debido a su arquitectura de codificador-decodificador y su enfoque text-to-text. Entre sus puntos fuertes del modelo se destacan: i) gracias al uso de atención cruzada (cross-attention), el modelo pudo integrar información del contexto con alta fidelidad, generando narrativas que reflejaron de manera precisa los datos de entrada, ii) su diseño permite adaptarse a tareas complejas, como resúmenes o comparaciones históricas, proporcionando flexibilidad en su aplicación, y iii) en las evaluaciones realizadas, T5 obtuvo mejores resultados en BLEU (0.14), BERTScore (0.83) y perplejidad, lo que evidencia su superioridad técnica en generación de textos estructurados. Por otra parte, en el modelo T5 se identificó que su arquitectura requiere más recursos computacionales para entrenamiento y generación, lo que podría limitar su uso en entornos con recursos restringidos.

En cuanto al modelo GPT-2, este mostró un buen desempeño en la generación de textos económicos, destacándose por su creatividad y variedad en las narrativas. Sus principales fortalezas fueron: i) generar textos coherentes y estilísticamente variados, lo que podría ser útil en aplicaciones que requieran creatividad y adaptabilidad y ii) su enfoque unidireccional le permite generar textos completos incluso con entradas limitadas. Por otra parte, en algunos casos, GPT-2

generó información irrelevante o incoherente, probablemente debido a su naturaleza autoregresiva. Aunque generó textos creativos, tuvo menor puntuación en métricas evaluadas.

En cuanto al proceso de optimización de parámetros del modelo, Los resultados obtenidos para ambos modelos revelan que la optimización de parámetros tuvo efectos diferentes en cada uno, pero en general, se observa una mejora en la calidad de las noticias generadas. En cuanto al modelo T5, este mostró una mejora moderada en las métricas de precisión y F1, aunque la perplejidad aumentó ligeramente. Esto sugiere que el modelo optimizado produce salidas con mayor precisión y relevancia, pero con una ligera mayor complejidad en la predicción de las palabras. Para el GPT-2, en cambio, se experimentó una mejora considerable en la perplejidad, lo que se traduce en una mayor fluidez y coherencia de las noticias generadas. Además, la mejora en el BLUE Score refleja una mayor alineación de las salidas generadas con las expectativas de los textos de referencia.

La comparación entre T5 y GPT-2 muestra cómo ambos modelos tienen aplicaciones complementarias dependiendo del contexto. En cuanto, a la precisión y la fidelidad, T5 demostró una mayor capacidad para generar textos alineados con los datos de entrada, superando a GPT-2 en todas las métricas evaluadas. En cuanto a la creatividad y la diversidad del texto, GPT-2 generó narrativas más variadas y con un estilo más libre, lo que lo hace útil en escenarios menos estructurados. Y en cuanto a la eficiencia computacional, GPT-2 requiere menos recursos computacionales, lo que podría ser ventajoso en implementaciones con limitaciones tecnológicas.

La optimización de parámetros mejoró significativamente la calidad de los textos generados. Por ejemplo, el ajuste de “temperature” y “top-p” permitió a ambos modelos equilibrar creatividad y precisión en las narrativas. En cuanto a las métricas aplicadas (BLEU, BERTScore, perplejidad) proporcionaron una evaluación cuantitativa del rendimiento, permitiendo identificar las fortalezas de cada modelo de manera clara. Y, mientras T5 se destacó en la generación de textos alineados con datos económicos, GPT-2 demostró ser más adecuado para narrativas generalistas o con menor necesidad de precisión absoluta.

8. Referencias Bibliográficas

[1] G. Bustamante, "Extracción de información para la generación de reportes estructurados a partir de noticias peruanas relacionadas a crímenes," *Pontif. Universdad Católica Perú*, no. 2, pp. 1–13, 2019.

[2] S. Crucianelli, "¿Qué es el periodismo de datos?," *Cuad. Periodis.*, vol. 26, pp. 106–124, 2013.

[3] R. García, "Desarrollo y uso de un software de generación de noticias deportivas con sentimiento en la mejora de procesos en el ámbito periodístico," *Univ. Carlos III*, 2020.

[4] Y. Ledeneva and R. García-Hernández, *Automatic Generation of Text Summaries: Challenges, proposals and experiments*. 2017.

[5] M. Elian and A. M. Becerra, "Una Revisión de la Generación Automática de Resúmenes Extractivos A Review of the Extractive Text Summarization," *UIS Ing. Rev. la Fac. Ing. Fis.*, no. l, pp. 7–27, 2012.

[6] A. Villa Monte, *Generación automática inteligente de resúmenes de textos con técnicas de Soft Computing*. 2019.

[7] N. Fatima, A. S. Imran, Z. Kastrati, S. M. Daudpota, and A. Soomro, "A Systematic Literature Review on Text Generation Using Deep Neural Network Models," *IEEE Access*, vol. 10, pp. 53490–53503, 2022, doi: 10.1109/ACCESS.2022.3174108.

[8] F. Kiyani and O. Tas, "A survey automatic text summarization," *Pressacademia*, vol. 5, no. 1, pp. 205–213, 2017, doi: 10.17261/pressacademia.2017.591.

[9] C. Schröer, F. Kruse, and J. M. Gómez, "A systematic literature review on applying CRISP-DM process model," *Procedia Comput. Sci.*, vol. 181, no. 2019, pp. 526–534, 2021, doi: 10.1016/j.procs.2021.01.199.

[10] N. Nazari and M. Mahdavi, "A survey on automatic text summarization techniques," *J. AI Data Min.*, vol. 2040, no. 1, pp. 121–135, 2019, doi: 10.1088/1742-6596/2040/1/012044.

[11] Y. J. Kumar, O. S. Goh, H. Basiron, N. H. Choon, and P. C. Suppiah, "A review on

automatic text summarization approaches," *J. Comput. Sci.*, vol. 12, no. 4, pp. 178–190, 2016, doi: 10.3844/jcssp.2016.178.190.

[12] A. Perez-Castro, M. R. Martínez-Torres, and S. L. Toral, "Efficiency of automatic text generators for online review content generation," *Technol. Forecast. Soc. Change*, vol. 189, no. January 2022, p. 122380, 2023, doi: 10.1016/j.techfore.2023.122380.

[13] A. P. Widyassari *et al.*, "Review of automatic text summarization techniques & methods," *J. King Saud Univ. - Comput. Inf. Sci.*, vol. 34, no. 4, pp. 1029–1046, 2022, doi: 10.1016/j.jksuci.2020.05.006.

[14] W. S. El-Kassas, C. R. Salama, A. A. Rafea, and H. K. Mohamed, "Automatic text summarization: A comprehensive survey," *Expert Syst. Appl.*, vol. 165, no. November 2021, 2021, doi: 10.1016/j.eswa.2020.113679.

[15] Amazon, "¿Qué es una RNN (red neuronal recurrente)?," 2024. .

[16] A. Vaswani *et al.*, "Attention Is All You Need Ashish," *31st Conf. Neural Inf. Process. Syst.*, no. Nips, pp. 4752–4758, 2017, doi: 10.1145/3583780.3615497.

[17] K. Kowsari, K. J. Meimandi, M. Heidarysafa, S. Mendu, L. Barnes, and D. Brown, "Text classification algorithms: A survey," *Inf.*, vol. 10, no. 4, pp. 1–68, 2019, doi: 10.3390/info10040150.