

Web Traffic Time Series Forecasting Using GRU Neural Networks with Sequence-to-Sequence Architecture: A Case Study at a National University

Gauto Sand Enrique¹, Rambo Alice Raquel¹

¹ Universidad Nacional de Misiones, Facultad de Ciencias Exactas Químicas y Naturales, Departamento de Informática, Misiones 3300 ARG

Abstract: This paper presents a case study on web traffic time series forecasting for the institutional websites of a national university in Argentina. The target variable considered in this study was the daily number of page views, obtained from historical Google Analytics records and complemented with features derived from users, sessions, acquisition channels, devices, and geographic origin. Data were collected from three institutional websites using both Universal Analytics (UA) and Google Analytics 4 (GA4). Gated Recurrent Unit (GRU) recurrent neural networks with a sequence-to-sequence (seq2seq) encoder-decoder architecture were implemented. The CRISP-DM methodology (Chapman et al., 2000; Schröder et al., 2021; Massahiro et al., 2024) guided the entire data mining process, including data collection, preprocessing, model design, training, and evaluation. The Hyperband algorithm was employed for hyperparameter optimization. Model performance was assessed using Mean Absolute Error (MAE), Root Mean Square Error (RMSE), and Symmetric Mean Absolute Percentage Error (SMAPE). Results show that, for the TUM Transmedia website, the best-performing model achieved an MAE of approximately 11 using UA data, outperforming comparable results reported in the literature. When applied to GA4 data, the model maintained an MAE below 30. For the Editorial Universitaria website, the results were less satisfactory, with an MAE of 71, reflecting the greater complexity and irregularity of its traffic dynamics. Furthermore, comparisons with baseline models revealed that GRU Seq2Seq outperformed classical forecasting approaches on relatively stable traffic patterns, while simpler methods such as Moving Average achieved better performance on highly volatile series. The study demonstrates the viability of GRU-based architectures for institutional web traffic forecasting and provides a foundation for future research on predictive analytics for university infrastructure management.

Keywords: web traffic forecasting, time series, GRU, recurrent neural networks, sequence-to-sequence

Pronóstico de Series de Tiempo de Tráfico Web Mediante Redes Neuronales GRU con Arquitectura Secuencia a Secuencia: Caso de Estudio en una Universidad Nacional

Resumen: Este artículo presenta un caso de estudio sobre el pronóstico de series de tiempo de tráfico web para las páginas institucionales de una universidad

nacional de Argentina. La variable objetivo considerada en este estudio fue el número diario de vistas de página (*page views*), obtenido a partir de registros históricos de Google Analytics, complementada con variables derivadas de usuarios, sesiones, canales, dispositivos y procedencia geográfica. Los datos provinieron de tres sitios web institucionales y fueron recolectados tanto mediante Universal Analytics (UA) como Google Analytics 4 (GA4), se implementaron redes neuronales recurrentes de Unidad Recurrente con Compuertas (GRU) con arquitectura secuencia a secuencia (seq2seq) codificador-decodificador. La metodología CRISP-DM (Chapman, et al., 2000; Schröer, et al., 2021; Massahiro, et al., 2024) guió el proceso completo de minería de datos, incluyendo recolección, preprocesamiento, diseño de modelos, entrenamiento y evaluación. Se utilizó el algoritmo Hyperband para la optimización de hiperparámetros. Los modelos fueron evaluados mediante el Error Medio Absoluto (MAE), la Raíz del Error Cuadrático Medio (RMSE) y el Error Absoluto Medio Porcentual Simétrico (SMAPE). Los resultados evidencian que, para el sitio TUM Transmedia, el mejor modelo alcanzó un MAE de aproximadamente 11 con datos de UA, superando resultados similares reportados en la literatura. Al aplicar el modelo sobre datos de GA4, se mantuvo un MAE inferior a 30. Para el sitio de la Editorial Universitaria, los resultados fueron menos satisfactorios, con un MAE de 71, lo que refleja la mayor complejidad de la dinámica de tráfico de ese sitio. El trabajo demuestra la viabilidad de las redes GRU para el pronóstico de tráfico web institucional y sienta las bases para futuras investigaciones sobre analítica predictiva en la gestión de infraestructura universitaria.

Palabras clave: pronóstico de tráfico web, series de tiempo, GRU, redes neuronales recurrentes, secuencia a secuencia

1. INTRODUCCIÓN

La gestión eficiente de la infraestructura tecnológica en instituciones de educación superior depende, en gran medida, de la capacidad de anticipar la demanda de recursos computacionales. En este contexto, el pronóstico del tráfico web de los sitios institucionales representa un insumo estratégico, ya que permite a los administradores de redes tomar decisiones informadas sobre la asignación y escalabilidad de recursos, previniendo situaciones como la caída de servidores, la lenta carga de páginas y la pérdida de usuarios (Kotu y Deshpande, 2019; Casado-Vara et al., 2021).

La institución objeto de este estudio ha registrado datos de tráfico web a través de la herramienta Google Analytics desde 2018, mediante Universal Analytics (UA). A partir de 2022, se incorporó también el registro a través de Google Analytics 4 (GA4). Sin embargo, según relevamientos realizados al personal técnico, estos datos prácticamente no habían sido utilizados con fines analíticos o predictivos. Esta situación pone en evidencia una oportunidad concreta para aplicar técnicas de minería de datos y aprendizaje automático sobre registros ya existentes.

En este trabajo, el problema de pronóstico se formuló como la predicción del número diario de vistas de página (*page views*) para sitios web institucionales. Esta variable constituye uno de los principales indicadores de demanda de recursos web y resulta relevante para la planificación de infraestructura tecnológica. Además de las vistas de página, se analizaron variables complementarias asociadas al comportamiento de los

usuarios y a las características de acceso registradas por Google Analytics. El objetivo general de este trabajo fue desarrollar y evaluar modelos de pronóstico de series de tiempo de tráfico web aplicados sobre datos históricos provenientes de las páginas institucionales, utilizando redes neuronales GRU con arquitectura secuencia a secuencia (seq2seq). La variable objetivo del modelo fue el número diario de vistas de página. Las demás variables utilizadas como entradas incluyeron usuarios, usuarios nuevos, sesiones, sesiones por usuario, país de origen, dispositivo, sistema operativo y canal de adquisición. Para abordar el problema se implementó una arquitectura GRU secuencia a secuencia (*seq2seq*) con esquema codificador-decodificador, orientada a pronosticar 60 días futuros a partir de ventanas históricas de 60 días. La arquitectura fue optimizada mediante Hyperband y evaluada utilizando MAE, RMSE y SMAPE.

2. REVISIÓN BIBLIOGRÁFICA

2.1. Series de tiempo y tráfico web

El pronóstico del tráfico web se realiza mediante el análisis de series de tiempo, un campo donde la literatura científica evalúa desde modelos estadísticos clásicos hasta arquitecturas avanzadas de aprendizaje profundo. Entre las soluciones más destacadas se encuentran la combinación de modelos generativos (GAN) con redes LSTM y MLP (Zhou et al., 2021), que toman como base la arquitectura original de memoria a corto y largo plazo (Hochreiter & Schmidhuber, 1997), así como el uso de redes LSTM con entrenamiento distribuido que logran un MAE inferior a 30 (Casado-Vara et al., 2021). Asimismo, se han desarrollado arquitecturas RNN seq2seq con codificador-decodificador GRU (Petluri & Al-Masri, 2018), modelos híbridos NARX-QPSO de inspiración cuántica, combinaciones Prophet-LGBM (Xu et al., 2021; Kuranga & Pillay, 2022), y el sistema automatizado Auto-PyTorch-TS (Deng et al., 2023). Finalmente, tecnologías como las redes LSTM bidireccionales (Bi-LSTM) han demostrado una alta precisión en el conjunto M3-Competition con un SMAPE de 8,55%, superando consistentemente a los métodos tradicionales como ARIMA (Quoc Nguyen, 2021).

2.2. Arquitectura GRU y seq2seq

Las redes GRU son una variante de redes neuronales recurrentes diseñadas para modelar datos secuenciales mediante mecanismos de actualización y reinicio del estado oculto. En comparación con LSTM, presentan menor complejidad computacional y suelen alcanzar desempeños comparables en tareas de predicción de secuencias. En este trabajo se utilizó una arquitectura GRU seq2seq codificador-decodificador, en la cual el codificador resume la ventana histórica de entrada y el decodificador genera el horizonte futuro de predicción.

2.3. Métricas de evaluación

Para la evaluación del desempeño de los modelos, la literatura especializada propone diversas métricas (Botchkarev, 2018; Wang y Lu, 2018). En el presente trabajo se emplearon las siguientes: el Error Medio Absoluto (MAE), que mide el promedio de

las diferencias absolutas entre los valores predichos y los reales; la Raíz del Error Cuadrático Medio (RMSE), que penaliza los errores grandes al elevar al cuadrado las diferencias; y el Error Absoluto Medio Porcentual Simétrico (SMAPE), que normaliza el error en términos porcentuales y resulta útil para comparar resultados entre distintos conjuntos de datos (Casado-Vara et al., 2021; Botchkarev, 2018).

2.4. Optimización de hiperparámetros

Los hiperparámetros son variables que se configuran antes del entrenamiento de un modelo y tienen un impacto directo tanto en su rendimiento como en su topología. En la tesis se utilizó el algoritmo Hyperband para su optimización, el cual es una extensión del algoritmo Successive Halving (Li, et al., 2018) que varía el número de configuraciones evaluadas y el presupuesto de recursos asignado a cada una, con el objetivo de encontrar eficientemente la mejor combinación posible. A diferencia de la búsqueda aleatoria (Bergstra, Yoshua, 2012) y la optimización bayesiana (Snoek, et al., 2012), Hyperband no requiere especificar un número máximo de pruebas (trials), sino únicamente el valor máximo de épocas y un factor de reducción, lo que lo hace especialmente eficiente en escenarios con recursos computacionales limitados.

En el trabajo, los hiperparámetros optimizados mediante *Hyperband* fueron: la cantidad de unidades GRU por capa (entre 128 y 512, con paso de 32), la tasa de aprendizaje del optimizador Adam (entre $1e-4$ y $1e-2$), el *batch size* (entre 64 y 256, con paso de 32) y el número de capas GRU (1 o 2). El algoritmo fue configurado con un objetivo de 200 épocas y un factor de reducción de 3. Cabe destacar que el número de N-días del codificador (fijado en 60) y el número de capas fueron determinados manualmente, ya que explorar más de dos capas con *Hyperband* agotaba rápidamente los recursos disponibles en Google Colab al generar cientos de modelos durante la búsqueda.

3. METODOLOGÍA Y DATOS

3.1. Marco metodológico

Se adoptó la metodología CRISP-DM, que comprende fases de comprensión del negocio, comprensión de los datos, preparación, modelado, evaluación y despliegue. El preprocesamiento de datos incluyó detección de anomalías (KNN) (Dang, et al., 2015), con $k=7$, normalización log1p y z-score, construcción de variables rezagadas y generación de estadísticas descriptivas. El modelado se realizó mediante una arquitectura GRU seq2seq, cuyos hiperparámetros fueron optimizados con Hyperband.

3.2. Conjuntos de datos

Se accedió a datos históricos de tráfico web de tres páginas institucionales: el sitio principal de la institución (unam.edu.ar), el sitio de la editorial universitaria (editorial.unam.edu.ar) y el portal de medios audiovisuales (transmedia.unam.edu.ar). Los datos de UA abarcaron el período comprendido entre mayo de 2018 y julio de 2023 aproximadamente, mientras que los datos de GA4 cubrieron desde junio de 2022 hasta

enero de 2024. La descarga de los datos se realizó conectando Google Analytics como fuente de datos en Looker Studio, exportando los registros en archivos CSV. Debido a un problema de calidad detectado, la ausencia de registros en GA4 para el sitio principal a partir de noviembre de 2023, ese conjunto de datos fue descartado. En consecuencia, los experimentos se realizaron sobre los conjuntos de datos correspondientes a TUM Transmedia y Editorial Universitaria, tanto para UA como para GA4.

Tabla 1. Detalle de los Sitios relevados, las fuentes de datos y los períodos analizados

Sitio	Fuente	Período
TUM Transmedia	UA	Ago. 2018 – Jun. 2023
Editorial Universitaria	UA	Oct. 2018 – Oct. 2023
TUM Transmedia	GA4	Jun. 2022 – Ene. 2024
Editorial Universitaria	GA4	Jun. 2022 – Ene. 2024

La variable objetivo considerada en todos los experimentos fue el número diario de vistas de página (page views). Como se puede observar en la Tabla 1 los conjuntos de datos fueron obtenidos a partir de Universal Analytics y Google Analytics 4 para los sitios TUM Transmedia y Editorial Universitaria. La utilización de ambas fuentes permitió evaluar el comportamiento de los modelos frente a la migración entre plataformas de analítica web. Inicialmente se analizaron tres sitios institucionales. Sin embargo, el sitio principal de la Universidad Nacional de Misiones fue excluido debido a problemas de calidad e inconsistencias detectadas durante la etapa de preparación de datos.

Como variables auxiliares se incorporaron atributos derivados de usuarios, sesiones, canales de adquisición, dispositivos, sistemas operativos y países de procedencia, así como variables temporales construidas mediante rezagos y medidas estadísticas descriptivas.

3.2.1 Análisis exploratorio de las series temporales

Se realizó un análisis exploratorio inicial sobre las series de vistas diarias correspondientes a los sitios institucionales estudiados. La Figura 1 presenta las series temporales de TUM Transmedia y Editorial Universitaria obtenidas a partir de Universal Analytics. Cada subgráfico se muestra con su propia escala vertical, debido a las diferencias de magnitud entre ambas series.

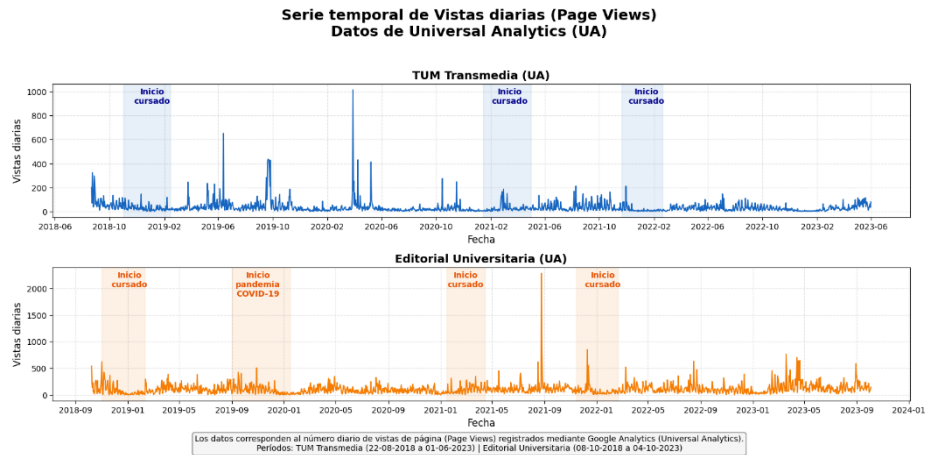


Figura 1 - Serie temporal de vistas diarias.

La Figura 1 presenta la evolución temporal del número diario de vistas de página (*page views*) registradas en los sitios TUM Transmedia y Editorial Universitaria utilizando datos de Universal Analytics. En ambas series se observa un comportamiento irregular, con variabilidad temporal marcada y presencia de picos abruptos de tráfico. En el caso de TUM Transmedia, la serie presenta picos aislados de gran magnitud, especialmente en ciertos períodos específicos, seguidos por intervalos de menor actividad. Por su parte, Editorial Universitaria exhibe una mayor amplitud relativa y una frecuencia más sostenida de episodios de alta volatilidad. Estas características sugieren la influencia de factores exógenos, tales como eventos institucionales, publicaciones específicas, campañas de difusión o actividades académicas, que afectan el comportamiento del tráfico web y aumentan la complejidad del problema de pronóstico.

Para complementar la inspección visual, se aplicaron las pruebas de estacionariedad Augmented Dickey-Fuller (ADF) y KPSS sobre la variable objetivo. Los resultados de la Tabla 2 muestran comportamientos heterogéneos entre las series: TUM GA4 presenta evidencia consistente de no estacionariedad, Editorial GA4 muestra un comportamiento compatible con estacionariedad, y las series provenientes de UA presentan resultados mixtos, lo que sugiere posibles tendencias, cambios de nivel o componentes estructurales.

En general se puede observar que TUM presenta picos aislados de gran magnitud, mientras que Editorial muestra una mayor frecuencia de episodios de alta volatilidad.

Tabla 2. Evaluación de estacionariedad de las series temporales mediante ADF y KPSS

Dataset	ADF p-value	KPSS p-value
TUM UA	< 0,001	0,010
Editorial UA	< 0,001	0,013
TUM GA4	0,253	0,010
Editorial GA4	< 0,001	0,100

3.3. Preprocesamiento de datos

El preprocesamiento incluyó la unificación de archivos CSV, limpieza de valores nulos, detección de outliers mediante KNN, normalización, generación de variables rezagadas y construcción de conjuntos de entrenamiento y prueba. Para la ingeniería de variables: se construyeron columnas derivadas como se observa en la Tabla 3, de las vistas por sistema operativo (Windows, Android, Linux, Macintosh, iOS, entre otros), vistas por agrupación de canales de tráfico (Organic Search, Social, Direct, Referral), vistas por categoría de dispositivo (Mobile, Desktop, Tablet) y vistas por país (Argentina y otros países). Además de las variables originales obtenidas desde Google Analytics, se construyeron atributos derivados con el objetivo de capturar patrones temporales y características del comportamiento de los usuarios. En particular, se incorporaron variables rezagadas (lags) de 90, 180, 270 y 360 días, adicionalmente, para cada agrupación se calcularon estadísticos descriptivos (media, desviación estándar y mediana). La incorporación de variables rezagadas permitió representar explícitamente patrones históricos de corto, mediano y largo plazo, complementando la capacidad de modelado temporal provista por la arquitectura GRU. También se utilizaron variables asociadas a dispositivos, sistemas operativos, canales de adquisición y procedencia geográfica de las visitas.

Tabla 3. Variables utilizadas durante el modelado

Grupo	Variables
Tráfico	vistas, usuarios, usuarios nuevos, sesiones
Dispositivo	mobile, desktop, tablet
Sistema operativo	Windows, Android, Linux, iOS, Macintosh
Canal	Organic Search, Social, Direct, Referral
Geografía	Argentina, otros países
Temporales	medianas, lag_90, lag_180, lag_270, lag_360

3.4. Partición de los datos

El conjunto total de datos fue dividido en subconjunto de entrenamiento y subconjunto de prueba. La partición se realizó respetando el orden temporal de las observaciones, de modo que el subconjunto de entrenamiento correspondió al período más antiguo y el subconjunto de prueba al más reciente, garantizando que el modelo fuera validado con datos no vistos durante el aprendizaje. Para evitar fuga de información (data leakage), los parámetros de normalización fueron calculados únicamente sobre el subconjunto de entrenamiento y posteriormente aplicados al subconjunto de prueba, preservando la independencia temporal de los datos utilizados para la evaluación.

3.5. Arquitectura del modelo

Los hiperparámetros optimizados mediante el algoritmo Hyperband fueron: número de unidades GRU (rango 128-512, paso 32), tasa de aprendizaje (rango $1e-4$ a $1e-2$), *batch size* (rango 64-256, paso 32) y número de capas (1 o 2). Se utilizó el optimizador Adam (Kingma, 2014). Hyperband fue configurado con 200 épocas como objetivo y un factor de reducción de 3. El número de N-días del codificador (60 días) fue fijado manualmente basándose en el código de referencia del ganador de la competencia Kaggle de pronóstico de tráfico web (Suilin, 2017), el cual fue adaptado para el presente trabajo.

3.6. Plan de pruebas

Se diseñaron ocho pruebas experimentales con el siguiente esquema: las pruebas 1 y 2 evaluaron distintas funciones de pérdida (MAE vs. Huber) con datos de TUM Transmedia UA; la prueba 3 exploró el efecto de aumentar el número de capas; las pruebas 4 y 5 aplicaron los modelos obtenidos a Editorial Universitaria UA; la prueba 6 transfirió el mejor modelo de TUM UA a datos de TUM GA4; y las pruebas 7 y 8 evaluaron los mejores modelos sobre datos de Editorial Universitaria GA4. El criterio de aceptación para un modelo satisfactorio fue establecido en $MAE < 30$, siguiendo el umbral reportado en la literatura comparable (Casado-Vara et al., 2021).

4. RESULTADOS

Se realizaron ocho pruebas principales sobre dos conjuntos de datos: TUM Transmedia y Editorial Universitaria.

- **TUM Transmedia (UA/GA4):** La Prueba 1 (con datos de UA) alcanzó un MAE de 11,12, superando los referentes bibliográficos. Al transferir el modelo a datos de GA4 (Prueba 6), se obtuvo un MAE de 26,54, manteniéndose por debajo del umbral de aceptabilidad de 30.
- **Editorial Universitaria:** Los resultados fueron menos precisos, con un MAE de 81,53 en la Prueba 5. Se observó que el modelo tiende a predecir la media de la serie, pero no logra capturar picos de volatilidad extrema.

4.1. Pruebas con datos de TUM Transmedia UA

La Prueba 1 (Configuración óptima): Evaluada con normalización $\log_{1p}$ y z-score (pérdida MAE). Obtuvo los mejores hiperparámetros (1 capa, 224 unidades GRU, learning rate: 0,0012, batch: 64) y rendimiento más alto (MAE: 11,13, SMAPE: 0,598, RMSE: 16,49). Esta configuración se fijó como base.

La Prueba 2 y 3 (Variantes): La prueba 2 (pérdida Huber) y la prueba 3 (2 capas en codificador/decodificador, 500 épocas) empeoraron el desempeño. La pérdida de eficiencia con una segunda capa se alineó con la literatura de referencia (Suilin, 2017), validando la elección de una sola capa.

4.2. Pruebas con datos de Editorial Universitaria UA

La Prueba 4 (Transferencia directa): El reentrenamiento del modelo óptimo de TUM UA durante 500 épocas falló notablemente (MAE: 258,67, RMSE: 364,64), mostrando un pronóstico creciente alejado de la realidad.

La Prueba 5 (Optimización): Se ejecutó una nueva búsqueda con Hyperband (160 unidades GRU, learning rate: 0,000134, batch: 192). Aunque mejoró sustancialmente (MAE: 81,54, SMAPE: 0,610, RMSE: 130,60), no logró bajar del umbral de aceptación establecido (MAE < 30).

4.3. Transferencia a datos de GA4

La Prueba 6 (TUM GA4): El modelo base de TUM UA reentrenado en TUM GA4 funcionó con éxito (MAE: 26,54, SMAPE: 0,579, RMSE: 34,66), superando el umbral de aceptación (MAE < 30) y demostrando su adaptabilidad a la nueva plataforma analítica.

La Pruebas 7 y 8 (Editorial GA4): Se evaluaron los modelos previos sobre los datos de Editorial GA4. La prueba 8 (modelo de Editorial UA) superó a la prueba 7 (modelo de TUM UA) con un MAE: 71,08 frente a 89,94. No obstante, ningún enfoque logró satisfacer el criterio de aceptación final (MAE < 30).

4.4. Análisis de MAE y RMSE en comparación con los enfoques de referencia.

Se adoptó MAE como métrica principal debido a que también fue utilizada como función de pérdida durante el entrenamiento. RMSE se empleó para penalizar errores grandes y SMAPE para comparar errores relativos entre series. Siguiendo la literatura, se consideró MAE < 30 como umbral de desempeño satisfactorio. Los resultados muestran que dicho umbral se alcanzó en TUM Transmedia, tanto en UA como en GA4, pero no en Editorial Universitaria.

La Tabla 4 resume los resultados de las ocho pruebas realizadas, incluyendo el tipo de comportamiento del pronóstico (media o tendencia) y los valores de las tres métricas de evaluación.

Tabla 4. Resumen de las pruebas realizadas

Prueba	Fuente	pronóstico	métrica MAE	métrica SMAPE	métrica RMSE
1	TUM UA	Media	11,13	0,598	16,49
2	TUM UA	Media	13,27	0,697	17,52
3	TUM UA (2 capas)	Tendencia	12,15	0,635	17,79
4	Editorial UA	Tendencia	258,67	1,184	364,64
5	Editorial UA	Media	81,54	0,610	130,60
6	TUM GA4	Tendencia	26,54	0,579	34,66
7	Editorial GA4 (mod. TUM UA)	Tendencia	89,94	0,957	129,76
8	Editorial GA4 (mod. Edit. UA)	Media	71,08	0,604	104,80

4.5 Comparación con modelos baseline

Con el objetivo de contextualizar el desempeño del modelo propuesto, se incorporaron tres modelos de referencia: Naïve, Media Móvil de 7 días y ARIMA. Los resultados obtenidos se presentan en la Tabla 5.

Tabla 5. Comparación GRU vs Naïve, Media Móvil y ARIMA

Dataset	Modelo	MAE	RMSE	SMAPE
TUM UA	Naïve	17,23	26,69	0,64
TUM UA	Media móvil 7 días	14,54	21,24	0,58
TUM UA	ARIMA(1, 1, 2)	19,51	25,48	0,78
TUM UA	GRU Seq2Seq	11,13	16,49	0,59
Editorial UA	Naïve	70,15	114,09	0,54
Editorial UA	Media móvil 7 días	65,79	97,52	0,51
Editorial UA	ARIMA(3, 0, 2)	72,06	111,04	0,58
Editorial UA	GRU Seq2Seq	81,54	130,60	0,61

Para el conjunto TUM UA, la arquitectura GRU Seq2Seq alcanzó el menor error absoluto medio (MAE = 11,13) y el menor error cuadrático medio (RMSE = 16,49), superando a los tres modelos baseline evaluados. Estos resultados sugieren que la arquitectura recurrente logra capturar dependencias temporales presentes en series con patrones relativamente regulares.

Sin embargo, para Editorial Universitaria UA los resultados fueron diferentes. En este caso, los modelos baseline obtuvieron errores inferiores a los alcanzados por la GRU Seq2Seq, siendo la Media Móvil de 7 días el método con mejor desempeño (MAE = 65,79 y RMSE = 97,53). Este comportamiento sugiere que la elevada volatilidad y la presencia de picos abruptos observados en la serie limitan la capacidad de generalización del modelo propuesto.

En conjunto, estos resultados indican que la efectividad de la arquitectura GRU Seq2Seq depende en gran medida de las características de la serie temporal analizada. Mientras que el enfoque mostró ventajas claras en TUM UA, no logró superar a modelos más simples en Editorial Universitaria UA.

5. DISCUSIÓN

Los resultados obtenidos para TUM Transmedia UA (MAE: 11,13) representan una mejora significativa respecto a los valores reportados por Casado-Vara et al. (2021), cuyo mejor resultado fue un MAE de 17,43 para el conjunto de datos de Wikipedia en idioma chino. Esta comparación es relevante dado que ambos analizan tráfico web y establecen un umbral de MAE < 30 como criterio de buen rendimiento.

La diferencia de rendimiento entre TUM Transmedia y Editorial Universitaria puede atribuirse a distintas dinámicas de los usuarios de cada sitio. TUM Transmedia, al ser un portal de medios audiovisuales, puede presentar patrones de acceso más regulares y predecibles, asociados a la periodicidad de sus contenidos. En contraste, la Editorial Universitaria podría atraer tráfico más irregular, influenciado por eventos específicos como lanzamientos de publicaciones o períodos académicos puntuales, generando picos difícilmente anticipables por el modelo. Este fenómeno es consistente con lo señalado por Casado-Vara et al. (2021), quienes indican que los modelos predictivos no son capaces de anticipar picos pronunciados asociados a comportamientos impredecibles de los usuarios.

El análisis de los comportamientos del pronóstico permitió identificar dos tipos de modelos: los modelos de media, que convergen al valor promedio de la serie a largo plazo con menor sensibilidad a la volatilidad, y los modelos de tendencia, que capturan la dirección y estacionalidad de la serie a costa de mayor error en los valores medios. En la práctica, la elección entre ambos tipos depende del objetivo de uso: para planificar el aprovisionamiento de recursos a largo plazo, un modelo de media puede resultar más conservador y seguro; mientras que para anticipar picos de demanda, un modelo de tendencia puede ser más informativo, aunque menos preciso en los valores intermedios. La transición de UA a GA4 fue una circunstancia no prevista que introdujo una variable adicional al estudio. El hecho de que el modelo entrenado con UA haya mantenido un $MAE < 30$ al ser reentrenado con datos de GA4 para TUM Transmedia sugiere una cierta robustez del modelo frente a cambios en la fuente de datos. Sin embargo, para la Editorial Universitaria no se logró la misma transferencia de rendimiento, lo que puede deberse a la menor cantidad de datos históricos disponibles en GA4, que limitó las posibilidades de entrenamiento.

Entre las limitaciones del presente trabajo se destacan: la restricción a un máximo de dos capas en el entrenamiento, condicionada por los recursos computacionales disponibles en Google Colab; la menor cantidad de datos en GA4 respecto a UA; la imposibilidad de anticipar picos esporádicos generados por comportamientos externos al modelo; y la exclusión de variables exclusivas de GA4 que no tienen correspondencia en UA y que podrían enriquecer la capacidad predictiva de los modelos. Estas limitaciones señalan direcciones claras para investigaciones futuras.

6. CONCLUSIONES

El presente trabajo desarrolló y evaluó modelos de pronóstico de series de tiempo de tráfico web aplicados a páginas institucionales de una universidad nacional, empleando redes neuronales recurrentes GRU con arquitectura secuencia a secuencia. A partir de la revisión del estado del arte, se identificó que esta arquitectura presenta resultados competitivos en la literatura, lo que motivó su selección e implementación siguiendo la metodología CRISP-DM.

Los experimentos realizados evidencian que el modelo propuesto es capaz de capturar la tendencia general del tráfico web en determinados contextos. Para el sitio TUM Transmedia, se obtuvieron modelos con MAE de 11,13 (UA) y 26,54 (GA4), ambos por debajo del umbral de aceptación de 30, constituyendo resultados competitivos respecto a los reportados en la literatura. Para el sitio de Editorial Universitaria, el mejor

resultado fue un MAE de 71, valor que no alcanzó el criterio de aceptación, posiblemente debido a la mayor irregularidad del tráfico y la menor cantidad de datos históricos disponibles.

El trabajo demuestra que los datos registrados por herramientas de analítica web, habitualmente subutilizados en las instituciones, constituyen un insumo valioso para el desarrollo de modelos predictivos que apoyen la gestión de infraestructura tecnológica. Asimismo, la transferencia de modelos entrenados con UA a datos de GA4 resultó viable para TUM Transmedia, sugiriendo que la arquitectura GRU seq2seq presenta cierta robustez ante cambios en la fuente de datos.

Los resultados obtenidos muestran que la arquitectura GRU Seq2Seq puede superar a modelos clásicos de pronóstico en conjuntos de datos con patrones relativamente estables, como TUM Transmedia. Sin embargo, en series caracterizadas por una mayor volatilidad y frecuencia de picos abruptos, como Editorial Universitaria, modelos más simples como la Media Móvil pueden alcanzar un desempeño comparable o superior. Estos hallazgos sugieren que la selección de la técnica de pronóstico debe considerar las características particulares de cada serie temporal y no únicamente la complejidad del modelo empleado.

Como futuras líneas de investigación se propone: desarrollar un sistema de pronóstico automatizado con actualización periódica de los modelos; incorporar variables exclusivas de GA4 no disponibles en UA; explorar arquitecturas de mayor profundidad con recursos computacionales más robustos; comparar el rendimiento de las redes GRU con otras arquitecturas de inteligencia artificial; y extender el análisis a otras páginas institucionales de la universidad una vez que se cuente con mayor cantidad de datos históricos en GA4.

REFERENCIAS BIBLIOGRÁFICAS

- Bergstra, J. y Yoshua, B. (2012). Random search for hyper-parameter optimization. *Journal of Machine Learning Research*, 13(10), 281-305.
- Blázquez-García, A., Conde, A., Mori, U. y Lozano, J. A. (2021). A review on outlier/anomaly detection in time series data. *ACM Computing Surveys*, 54(3).
- Botchkarev, A. (2018). Performance metrics (error measures) in machine learning regression, forecasting and prognostics: Properties and typology. <https://arxiv.org/abs/1809.03006>
- Casado-Vara, R., del Rey, A. M., Pérez-Palau, D., De-La-fuente-valentín, L. y Corchado, J. M. (2021). Web traffic time series forecasting using LSTM neural networks with distributed asynchronous training. *Mathematics*, 9(4), 421.
- Chapman, P., Clinton, J., Kerber, R., Khabaza, T., Reinartz, T., Shearer, C. y Wirth, R. (2000). CRISP-DM 1.0: Step-by-step data mining guide. SPSS Inc., 9(13), 1-73.
- Cho, K., van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H. y Bengio, Y. (2014). Learning phrase representations using RNN encoder-decoder for statistical machine translation. En *Proceedings of the 2014 Conference on Empirical Methods in Natural Language Processing (EMNLP)* (pp. 1724-1734).
- Chollet, F. K. (2015). Available online: <https://keras.io> (accessed on 14 August 2019). © 2019 by the authors. Licensee MDPI, Basel, Switzerland. This Article is an open access Article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).
- Dang, T. T., Ngan, H. Y. T. y Liu, W. (2015). Distance-based k-nearest neighbors outlier detection method in large-scale traffic data. *International Conference on Digital Signal Processing*, 2015, 507-510.

Deng, D., Karl, F., Hutter, F., Bischl, B. y Lindauer, M. (2023). Efficient automated deep learning for time series forecasting. En *Lecture Notes in Computer Science*, 13715, 664-680.

Hochreiter, S. y Schmidhuber, J. (1997). Long short-term memory. *Neural Computation*, 9(8), 1735-1780.

Kingma, D. P. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kotu, V. y Deshpande, B. (2019). Chapter 12 - Time series forecasting. En *Data Science (Second Edition)* (pp. 395-445). Morgan Kaufmann.

Kuranga, C. y Pillay, N. (2022). A comparative study of nonlinear regression and autoregressive techniques in hybrid with particle swarm optimization for time-series forecasting. *Expert Systems with Applications*, 190.

Li, L., Jamieson, K., DeSalvo, G., Rostamizadeh, A. y Talwalkar, A. (2018). Hyperband: A novel bandit-based approach to hyperparameter optimization. *Journal of Machine Learning Research*, 18, 1-52.

Massahiro, A., Cordeiro, R. F. y Goldman, A. (2024). The evolution of CRISP-DM for data science: Methods, processes and frameworks. *SBC Computing Review*, 28-43.

Nguyen, D. Q., Phan, M. N. y Zelinka, I. (2021). Periodic time series forecasting with bidirectional long short-term memory. *ACM International Conference Proceeding Series*, 60-64.

Petluri, N. y Al-Masri, E. (2019). Web traffic prediction of Wikipedia pages. En *Proceedings of the 2018 IEEE International Conference on Big Data* (pp. 5427-5429).

Schröer, C., Kruse, F. y Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. *Procedia Computer Science*, 181, 526-534.

Snoek, J., Larochelle, H. y Adams, R. P. (2012). Practical Bayesian optimization of machine learning algorithms. *Advances in Neural Information Processing Systems*, 25.

Suulin, A. (2017). kaggle-web-traffic [Código fuente]. GitHub. <https://github.com/Arturus/kaggle-web-traffic>

Sutskever, I., Vinyals, O. y Le, Q. V. (2014). Sequence to sequence learning with neural networks. *Advances in Neural Information Processing Systems*, 4, 3104-3112.

Wang, W. y Lu, Y. (2018). Analysis of the mean absolute error (MAE) and the root mean square error (RMSE) in assessing rounding model. *IOP Conference Series: Materials Science and Engineering*, 324(1).

Xu, S., Han, C. y Ran, C. (2021). A time series combined forecasting model based on Prophet-LGBM. *ACM International Conference Proceeding Series*.

Zhou, K., Wang, W., Huang, L. y Liu, B. (2021). Comparative study on the time series forecasting of web traffic based on statistical model and generative adversarial model. *Knowledge-Based Systems*, 213, 106467.