Georgia Tech Panama
Logistics Innovation & Research Center

Nota Técnica

Proyecciones de Flujos Logísticos

Preparado por: Mungyu Yang

Centro de Innovación e Investigaciones Logísticas Georgia Tech Panamá

Junio, 2025

www.gatech.pa | logistics.gatech.pa

Contenido

Resumen Ejecutivo

Este documento técnico ha sido desarrollado por el Centro de Innovación e Investigaciones Logísticas Georgia Tech Panamá con el objetivo de servir como guía metodológica para la implementación de una plataforma de análisis y proyección del desempeño logístico nacional, con énfasis en el tráfico portuario de contenedores. Está dirigido a analistas, técnicos, tomadores de decisiones y actores del ecosistema logístico interesados en comprender el funcionamiento interno del sistema y la base estadística de sus resultados.

El documento describe detalladamente el proceso seguido para integrar datos logísticos y económicos de diversas fuentes —puertos, comercio exterior, transporte terrestre—, y explica las metodologías utilizadas para su análisis. Incluye desde la selección de variables y pruebas de estacionariedad hasta técnicas de correlación y modelos de predicción basados en algoritmos de machine learning y redes neuronales

Asimismo, se explican los enfoques de validación (como la validación cruzada con ventana expandible), la selección de hiperparámetros , las métricas de evaluación y el cálculo de importancia de variables mediante técnicas como permutation importance. También se documenta el proceso de proyección futura con variables exógenas pronosticadas, asegurando un entorno de simulación realista y replicable.

En conjunto, esta guía busca proporcionar transparencia sobre el funcionamiento interno de la plataforma, estandarizar las prácticas analíticas utilizadas. Su propósito es fortalecer las capacidades analíticas del sector logístico panameño, promoviendo el uso riguroso de datos y modelos como insumo clave para la toma de decisiones estratégicas basadas en evidencia.

1) Introducción

El Centro de Innovación e Investigaciones Logísticas Georgia Tech Panamá ha desarrollado una plataforma centralizada y automatizada que integra datos logísticos de sectores como puertos, comercio exterior y transporte terrestre. A partir de este esfuerzo, se impulsa un nuevo proyecto piloto enfocado en analizar las interrelaciones entre estas variables y proyectar el desempeño logístico nacional. El piloto se centra en la logística marítima, que representa cerca del **30 % del Producto Interno Bruto** y posiciona a Panamá como líder regional en conectividad portuaria **(UNCTAD, 2023)**. Aunque el país ocupa una posición estratégica en el comercio global, los datos portuarios han estado fragmentados y subutilizados. **Esta plataforma permite proyectar tendencias en el movimiento portuario, identificar patrones macro-logísticos y apoyar la toma de decisiones en los sectores público y privado.**

1.1 Datos y variables

1.1.1 Variable dependiente

- Este proyecto utiliza el tráfico marítimo/movimiento de contenedores en los cinco principales puertos de Panamá como variable dependiente.
  - Representa **un indicador clave** del desempeño económico y logístico marítimo del país.
  - Aunque existe abundante literatura sobre análisis y predicción del tráfico portuario, son escasos los estudios en el contexto de América Latina.
- El dato abarca el período desde octubre de 2016, seleccionado estratégicamente tras la ampliación del Canal de Panamá, momento a partir del cual el movimiento de carga experimentó un crecimiento exponencial.

- El estudio se centra en dos tipos principales de movimiento de contenedores en Panamá:
  - **El trasbordo**, que representa aproximadamente el 90 % del total
    - En este proyecto, el trasbordo se analiza de forma diferenciada entre las dos costas del país:
      - **Costa Atlántica**, que incluye los puertos de Manzanillo, Colón y Cristóbal
      - **Costa Pacífica**, conformada por los puertos de Balboa y Rodman (PSA).
  - **El tráfico local**, correspondiente al 10 % restante, el cual abarca las importaciones y exportaciones panameñas. Esta proporción se obtiene restando del movimiento total de contenedores en Panamá el total asociado al trasbordo.

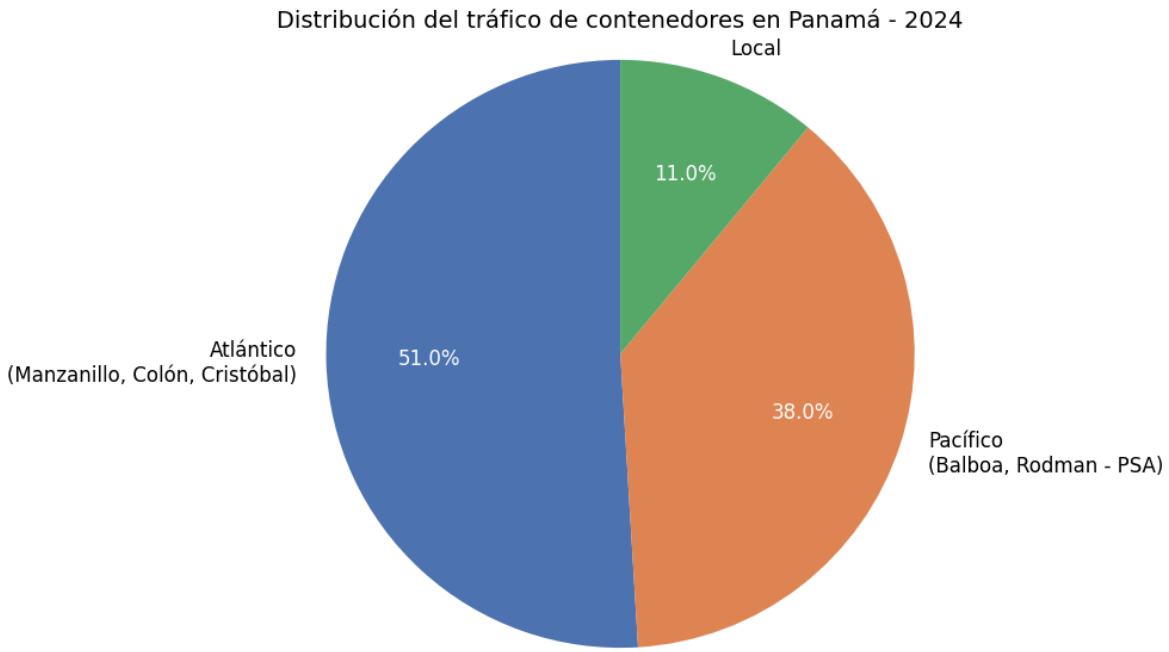

Ilustración 1: Distribución porcentual del tráfico de contenedores en los principales puertos de Panamá durante el año 2024, según tipo de operación: transbordo Atlántico, transbordo Pacífico y tráfico local. Fuente: Elaboración propia, datos de Autoridad Marítima de Panamá.

1.1.2 Variable independiente

- Este proyecto utiliza variables independientes que guardan relación con las variables dependientes.

- Criterio de selección de variables

    - **Revisión de literatura académica reciente**
        - Se consultan estudios recientes sobre proyección de tráfico portuario y comercio internacional.

        - Se identifican variables comúnmente utilizadas como predictores en modelos estadísticos y de machine learning.

        - Se priorizan aquellos enfoques aplicados en contextos similares al sistema portuario de Panamá.

    - **Conocimiento técnico del sistema logístico panameño**
        - Se incorporan variables específicas del entorno logístico nacional (capacidad portuaria, conectividad marítima, indicadores de servicios).

        - Se consideran dinámicas operativas diferenciadas entre el Atlántico y el Pacífico.

        - Aporta una perspectiva contextual y práctica sobre la influencia de ciertos indicadores locales en el movimiento de contenedores.

    - **Relevancia económica y geográfica del país**
        - Se seleccionan variables que reflejan la interacción económica de Panamá con socios estratégicos (EE. UU., China, Países Bajos, etc.).

        - Incluye indicadores macroeconómicos (producción manufacturera, comercio exterior, etc.) que impactan la demanda logística regional.

        - Permite capturar relaciones estructurales entre la economía global y el hub logístico panameño.

- Las variables de trasbordo dependen principalmente **de factores externos** como el comercio EE.UU.–China, mientras que las de tráfico local responden principalmente **a factores internos** como el comercio en zonas francas.

    - **Ejemplos**
        - **Portuarias**: Conectividad y capacidad de servicios (Balboa, Manzanillo, Colón, Rodman, Cristóbal)

        - **Comercio exterior**: Importaciones/exportaciones de EE.UU., Países Bajos, México, China y zonas francas panameñas (ZLC, Panamá Pacífico)

        - **Variables macroeconómicas**: Índices manufactureros (ISM), precios de importación, flujos fronterizos (EE.UU.-México)

2) Análisis

La plataforma aplica metodologías validadas en estudios recientes del ámbito logístico y portuario para analizar relaciones entre variables clave. Se integran dos enfoques principales: el análisis de series temporales y la correlación de Pearson.

Primero, a través del análisis de series temporales, los usuarios pueden observar cómo evolucionan las variables a lo largo del tiempo, identificar patrones recurrentes, y detectar tendencias o desviaciones relevantes. Esta visualización dinámica facilita una comprensión preliminar del comportamiento de las variables logísticas y económicas.

Posteriormente, se aplica la correlación de Pearson, una medida estadística ampliamente utilizada para cuantificar la fuerza y dirección de la relación lineal entre dos variables. Para asegurar la validez de los resultados, esta metodología se limita a variables que cumplen con criterios de estacionalidad y estacionalidad, verificados mediante pruebas como la prueba de Dickey-Fuller Aumentada (ADF).

Además, se calcula el valor p asociado a cada coeficiente de correlación, lo que permite evaluar su significancia estadística. Esto es fundamental para filtrar relaciones espurias, es decir, aquellas correlaciones aparentes que no reflejan una relación causal real entre las variables.

Este enfoque combinado de análisis temporal y validación estadística permite identificar conexiones relevantes y robustas entre variables, sirviendo como base confiable para la toma de decisiones estratégicas en el ámbito logístico.

2.1 Comparación de series temporales

- La comparación de series temporales constituye una herramienta metodológica fundamental para analizar relaciones dinámicas

entre variables logísticas y económicas. Al alinear uno o dos series, se busca identificar patrones de comportamiento conjunto, como sincronías, rezagos estructurados (lags), o rupturas de tendencia, que puedan indicar una posible interdependencia o relación causal.

- Esta metodología permite, por ejemplo, evaluar si los cambios en una variable explicativa —como el volumen de comercio exterior o los precios de insumos— anteceden o evolucionan con el desempeño portuario medido en unidades TEU. A través del análisis conjunto de estas series, es posible detectar correlaciones temporales, efectos retardados y relaciones no evidentes en análisis estáticos.

2.2 Análisis de correlación Pearson

2.2.1 Metodología

En la plataforma, se implementa un enfoque estadístico riguroso para identificar relaciones lineales entre variables económicas y logísticas mediante el uso del coeficiente de correlación de Pearson. Esta técnica permite cuantificar la intensidad y dirección de la relación lineal entre dos series temporales, proporcionando un valor entre -1 y 1 que indica si la asociación es negativa, positiva o inexistente. Como parte del flujo analítico, las series temporales se visualizan inicialmente en su forma original (valores crudos), lo que facilita al usuario detectar tendencias, patrones compartidos o posibles sincronías de forma intuitiva.

Sin embargo, dado que muchas de estas series presentan comportamientos no estacionarios —como tendencias de crecimiento, ciclos o estacionalidades—, es necesario aplicar un proceso de validación estadística antes de calcular la correlación. Por esta razón, la plataforma incorpora pruebas formales de estacionalidad como la prueba de Dickey-Fuller Aumentada (ADF), y transforma las variables cuando es necesario mediante diferenciación, asegurando que las correlaciones obtenidas

sean estadísticamente sólidas y confiables. Esta práctica es consistente con hallazgos recientes en la literatura, que demuestran que el uso del coeficiente de correlación de Pearson es válido **únicamente cuando al menos una de las series es estrictamente estacionaria**, como lo evidencia el trabajo de Yuan y Shou (2024), quienes también aplican ADF antes de calcular correlaciones para evitar resultados espurios.

Pasos metodológicos para el análisis de correlación con Pearson

1. **Exploración inicial con valores crudos:**
   Las variables se visualizan sin transformar para permitir una exploración visual de tendencias, ciclos y relaciones aparentes entre las series logísticas y económicas.

2. **Verificación de estacionalidad (Prueba ADF):**
   Se aplica la prueba de Dickey-Fuller Aumentada (ADF) para evaluar si cada serie mantiene una media y varianza constantes en el tiempo. Si no lo hace, se considera no estacionaria.

3. **Transformación por diferenciación:**
   Las series no estacionarias se transforman aplicando diferencias sucesivas (primera o segunda) hasta alcanzar estacionalidad, eliminando así efectos de tendencia o estacionalidad.

4. **Cálculo del coeficiente de correlación de Pearson:**
   Una vez transformadas, se calcula la correlación de Pearson entre las dos series estacionarias para medir su relación lineal.

5. **Evaluación de significancia estadística (valor p):**
   Se calcula el valor p del coeficiente para determinar si la relación observada es estadísticamente significativa, bajo la hipótesis nula de no correlación.

6. **Filtrado de relaciones espurias:**
   Solo se consideran aquellas correlaciones con valor p significativo.

Esto permite descartar asociaciones que podrían ser producto del azar o de estructuras comunes no causales.

3) Proyección

Los modelos de proyección implementados en la plataforma fueron seleccionados con base en estudios recientes y adaptados al contexto operativo portuario de Panamá. Se incluyeron arquitecturas neuronales como **LSTM y GRU**, especializadas en capturar dinámicas secuenciales no lineales, así como modelos tradicionales de tipo aprendizaje automático como **Random Forest**.

La validación se diseñó respetando el carácter temporal de los datos: las redes neuronales fueron evaluadas mediante validación cruzada con ventana expandible (expanding window), mientras que los modelos clásicos utilizaron una división cronológica (75 % entrenamiento, 25 % validación). Solo se consideraron los **tres modelos con mejor desempeño validado para cada variable dependiente** (menor RMSE y mayor R^2) para mostrar resultados en la plataforma.

Las proyecciones finales fueron generadas tras reentrenar los modelos sobre todo el historial disponible y utilizando **variables exógenas pronosticadas previamente**. Este enfoque garantiza que los resultados reflejen un entorno de predicción realista y validado, maximizando la utilidad de las estimaciones para la planificación logística y de infraestructura.

A continuación, se describen las etapas seguidas en el proceso de modelado.

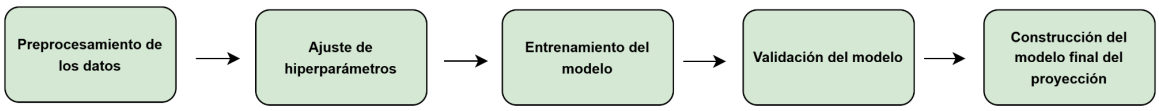

Ilustración 2: las etapas seguidas en el proceso de modelado. Fuente: Elaboración propia.

3.1 Modelos implementados

3.1.1 Long Short-Term Memory (LSTM)

Descripción: LSTM (Long Short-Term Memory) es una red neuronal recurrente especializada en el modelado de secuencias largas. Su arquitectura incorpora compuertas de entrada, olvido y salida que permiten retener información relevante durante periodos prolongados, resolviendo los problemas de desvanecimiento del gradiente presentes en redes RNN tradicionales. En el ámbito de la proyección portuaria, LSTM es especialmente útil para capturar dependencias temporales de largo plazo, patrones irregulares y relaciones complejas entre múltiples variables exógenas, como demanda logística, comercio exterior o condiciones operativas.

Referencia: En el estudio *A Hybrid Deep Learning Model for Port Throughput Forecasting: LSTM Combined with Empirical Mode Decomposition* (Wang et al., 2022), se aplicó un modelo LSTM híbrido al Puerto de Shenzhen. El modelo LSTM, incluso sin descomposición previa, logró un desempeño competitivo, alcanzando un RMSE de 41.82 y un R^2 de 0.977, superando a enfoques clásicos como ARIMA y SVM. Su capacidad para modelar series no lineales y tendencias multiescalares lo posiciona como una herramienta robusta para el análisis portuario.

3.1.2 Gated Recurrent Unit (GRU)

Descripción: GRU (Gated Recurrent Unit) es una red neuronal recurrente diseñada para modelar secuencias temporales con eficiencia computacional mejorada respecto a LSTM. Su arquitectura simplifica el mecanismo de puertas al combinar la compuerta de olvido y de entrada en una sola compuerta de actualización, lo cual reduce la complejidad del modelo y acelera el entrenamiento. En el contexto de la proyección de carga portuaria, GRU ha demostrado ser eficaz en la captura de patrones

no lineales y efectos estacionales en series de tiempo con múltiples factores externos.

Referencia: En el estudio de Chen y Huang (2020), se propuso un modelo GRU (Gated Recurrent Unit) optimizado con Adam para predecir la carga del Puerto de Guangzhou. Este modelo superó a BP (Backpropagation Neural Network), RNN (Recurrent Neural Network) y LSTM (Long Short-Term Memory) en todos los indicadores, logrando un RMSE de 59.138 y un R^2 de 0.973. También destacó por su rapidez de entrenamiento y menor número de ciclos, lo que lo hace ideal para contextos con datos operativos limitados.

3.1.3 Random Forest

Descripción: Random Forest es un algoritmo de aprendizaje supervisado por ensamble que combina múltiples árboles de decisión para mejorar la precisión y estabilidad de las predicciones. En el ámbito de proyección del rendimiento portuario, Random Forest ha demostrado un desempeño sobresaliente al modelar relaciones complejas y no lineales entre múltiples indicadores operativos del puerto, como el tiempo de estadía del buque, la productividad de muelle y la profundidad del calado.

Referencia: En el estudio de Awah, Nam y Kim (2021), el modelo Random Forest obtuvo el mejor desempeño entre siete métodos, con un RMSE de 0.0576, destacándose como el más preciso para estimar el rendimiento del puerto de Douala. Además, su interpretabilidad mediante importancia por permutación y gráficos de dependencia parcial lo convierte en una herramienta útil para la planificación y mejora operativa.

3.1.4 Regresión de Lasso y Ridge

Descripción: Lasso (Least Absolute Shrinkage and Selection Operator) y Ridge son modelos de regresión lineal regularizada ampliamente utilizados en problemas de predicción multivariada. Ambos incorporan penalizaciones para reducir el sobreajuste: Lasso aplica una penalización

L1 que tiende a seleccionar automáticamente las variables más relevantes, mientras que Ridge utiliza una penalización L2 que reduce el impacto de colinealidades manteniendo todos los predictores. En el ámbito portuario, estas técnicas son particularmente útiles para pronosticar indicadores como el volumen de contenedores, al permitir un manejo eficiente de series temporales con múltiples variables correlacionadas y un número limitado de observaciones.

Referencia: En el estudio de Liu Yuan (2020), ambos modelos obtuvieron los mejores resultados al predecir los volúmenes de contenedores vacíos en los puertos de Long Beach y Los Ángeles entre 1995 y 2009. LASSO y Ridge superaron en precisión a modelos como KNN, Random Forest y Gradient Boosting, con errores medios absolutos (MAE) de aproximadamente 10,300 TEUs y un desempeño cercano al método estadístico Winters. Además, el análisis de pruebas t pareadas mostró que no había diferencia estadísticamente significativa entre ambos, consolidándolos como modelos base eficaces para la proyección portuaria.

3.1.5 XGBoost

Descripción: XGBoost (Extreme Gradient Boosting) es un algoritmo de aprendizaje supervisado basado en árboles de decisión que utiliza una estrategia de ensamble de boosting para mejorar progresivamente la precisión de las predicciones. A diferencia de los modelos tradicionales de series temporales o redes neuronales complejas como LSTM o GRU, XGBoost destaca por su capacidad de manejar datos no lineales, trabajar con datos faltantes, evitar el sobreajuste mediante regularización y ofrecer una interpretación clara a través del análisis de importancia de variables. Estas características lo convierten en una opción ideal para la proyección de series temporales como el volumen de carga portuaria.

Referencia: El estudio de Nguyen y Cho (2024) reporta que el modelo XGBoost logró un error porcentual absoluto medio (MAPE) de tan solo 4.3%, superando ampliamente a SARIMA (22.14%), LSTM (18.10%) y modelos basados en descomposición de series temporales (10.95%). Además, XGBoost permitió identificar meses con desviaciones significativas, lo que proporciona a los responsables de la toma de decisiones una herramienta robusta para anticipar patrones irregulares relacionados con factores externos como la recesión económica global o la inflación.

3.1.6 SARIMAX

Descripción: SARIMAX (Seasonal AutoRegressive Integrated Moving Average with eXogenous regressors) es un modelo estadístico ampliamente utilizado para el análisis y pronóstico de series temporales que presentan componentes estacionales y relaciones con variables externas. Combina la capacidad de modelar tendencias y estacionalidad (propia del modelo SARIMA) con la incorporación de regresores exógenos, lo que permite capturar el impacto de factores externos en la variable objetivo. Es particularmente útil en contextos logísticos y portuarios donde las decisiones operativas dependen de condiciones macroeconómicas, estacionales y eventos exógenos.

Referencia: En el estudio de Lee y Bang (2024), el modelo SARIMAX se utilizó para pronosticar el movimiento de contenedores en el Puerto de Singapur, incorporando variables exógenas como el precio del petróleo WTI y las exportaciones de China. El modelo superó en precisión a métodos como LSTM, XGBoost y Prophet, demostrando que, al integrar factores externos clave, SARIMAX ofrece una solución eficaz y explicable para la previsión portuaria en contextos de alta estacionalidad.

3.2 Estrategia de ajuste de hiperparámetros y validación

La validación de modelos es esencial para evaluar qué tan bien un modelo de proyección generaliza a datos no vistos. A diferencia del aprendizaje automático tradicional —donde se suelen mezclar los datos aleatoriamente— en series temporales debemos respetar el orden cronológico de los datos para evitar filtraciones de información futura.

3.2.1 Ajuste de hiperparámetros con ventana expandible (Expanding Window Cross-Validation)

Este enfoque simula condiciones reales de proyección y utiliza una ventana expandible (Expanding Window Cross-Validation). Este tipo de validación para series temporales entrena siempre con datos previos y valida con datos posteriores, garantizando que no se utilice información futura, lo que coincide con lo descrito por Bergmeir y Benítez (2012), quienes destacan la importancia de respetar la cronología para evitar sesgos en la estimación del desempeño.

¿Cómo funciona?

- Se define una división inicial del conjunto de datos, reservando el 75 % de las observaciones para entrenamiento y el 25 % para validación.
- Sobre esta división se aplica TimeSeriesSplit, una técnica de validación que preserva el orden cronológico de los datos, asegurando que la información futura no sea utilizada durante el entrenamiento.
- Se realiza una búsqueda de hiperparámetros GridSearchCV) para cada modelo, optimizando el desempeño en la ventana de validación mediante la métrica de error cuadrático medio negativo.

- Una vez identificado el mejor conjunto de parámetros (por ejemplo, alpha para Lasso o Ridge), se entrena el modelo final con todos los datos disponibles.

- Finalmente, se utiliza el modelo ajustado para generar proyecciones realistas del rendimiento portuario en un horizonte de 12 meses, empleando como insumo las variables exógenas también pronosticadas.

¿Por qué se usa este método?

- Reproduce condiciones reales de proyección en las que el modelo no tiene acceso a información futura.

- Aumenta la robustez del proceso de selección de hiperparámetros al validar el modelo en múltiples particiones temporales.

- Minimiza el riesgo de sobreajuste y garantiza una evaluación honesta de la capacidad predictiva del modelo.

- Facilita la comparación objetiva entre algoritmos clásicos (Ridge, Lasso) y modelos más complejos como Random Forest

3.2.2 Ajuste de hiperparámetros con KerasTuner

El rendimiento de modelos como LSTM o GRU depende en gran medida de la correcta selección de hiperparámetros, como el número de capas, unidades por capa, *dropout*, tasa de aprendizaje y funciones de activación. Para evitar una selección manual —que requiere experiencia avanzada— se utilizó la biblioteca **KerasTuner**, que permite automatizar la búsqueda de combinaciones óptimas.

Este enfoque es útil para encontrar configuraciones base de forma eficiente. No obstante, estudios recientes (Shawki et al., 2021) señalan que, en problemas reales complejos, puede ser insuficiente sin una validación cuidadosa del espacio de búsqueda.

¿Cómo funciona?

- Se define un espacio de búsqueda con rangos para cada hiperparámetro relevante.
- El buscador (RandomSerach) de KerasTuner explora combinaciones de estos valores para encontrar la configuración que minimiza el error en los datos de validación.
- Para cada combinación, el modelo se entrena desde cero y se evalúa sobre un conjunto separado de datos (no usado en el entrenamiento), garantizando una comparación justa.
- Se utiliza la pérdida de error cuadrático medio (*MSE*) como métrica de evaluación durante el proceso de ajuste.

¿Por qué se usa este método?

- Permite encontrar automáticamente la arquitectura óptima para el modelo LSTM sin necesidad de prueba y error manual.
- Mejora la capacidad de generalización del modelo, ya que evita el sobreajuste a configuraciones arbitrarias.
- Aumenta la eficiencia del entrenamiento, encontrando una arquitectura balanceada entre complejidad y precisión.

Aplicación en el pipeline

En este proyecto, el ajuste automático se aplica principalmente al modelo LSTM, mientras que el modelo GRU se entrena con una arquitectura predefinida, basada en configuraciones óptimas obtenidas previamente. El modelo resultante con los mejores hiperparámetros se entrena posteriormente sobre todos los datos disponibles antes de realizar predicciones futuras.

Este enfoque asegura que el modelo utilizado para las proyecciones haya sido previamente validado y optimizado sobre datos reales, maximizando así la confiabilidad de los resultados proyectados.

3.2.3 División de entrenamiento y validación

Para los modelos no secuenciales, se aplicó una estrategia de validación basada en una división simple pero efectiva de los datos, respetando estrictamente el orden cronológico de la serie temporal. Esto asegura una evaluación realista del desempeño predictivo del modelo ante datos verdaderamente no vistos.

¿Cómo funciona?

- Se utiliza aproximadamente el **75 %** de los datos iniciales para entrenamiento.
- El **25 % restante** se reserva para validación final.
- El modelo se entrena exclusivamente con los datos del conjunto de entrenamiento.
- Luego, se evalúa su capacidad para predecir correctamente los valores del período reservado.

¿Por qué se usa este método?

- Es adecuado para modelos como **Random Forest o regresión Ridge**, que no se ajustan iterativamente con nuevas observaciones.
- Permite **evaluar modelos GRU y LSTM** en contextos reales, con la secuencia temporal conservada.
- Requiere menos recursos computacionales en comparación con validaciones cruzadas más complejas.

- Es interpretativamente más clara para decisiones operativas o políticas, pues simula un escenario real de predicción futura.

Ejemplo

- A continuación, se presenta la validación del modelo GRU aplicado a la predicción mensual del movimiento de contenedores en los **tres principales puertos atlánticos de Panamá (Puerto de Manzanillo, Puerto de Colón y Puerto de Cristóbal)**. La línea azul representa los valores reales observados y la línea naranja punteada representa las predicciones del modelo sobre el 25 % reservado para validación. El período de prueba abarca desde febrero de 2023 hasta marzo de 2025.

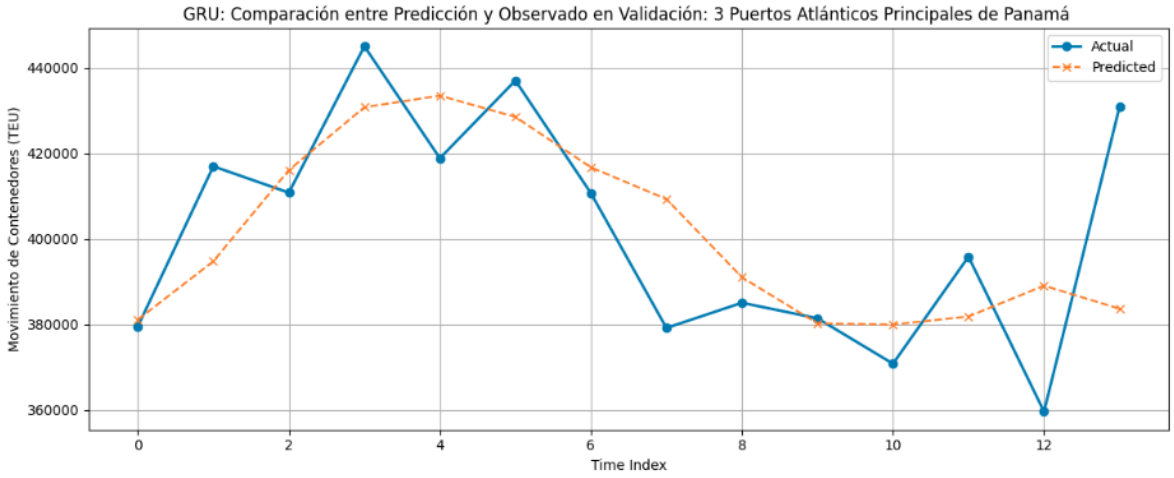

Ilustración 3: Comparación entre valores observados y predichos mediante un modelo GRU (Gated Recurrent Unit) durante la fase de validación para los tres principales puertos atlánticos de Panamá. Fuente: Elaboración propia.

3.3 Métricas de evaluación

La evaluación de modelos de proyecciónno se limita únicamente a qué tan cercanas son las predicciones a los valores reales. También debe considerar la consistencia, la dirección de los cambios y la robustez del modelo a lo largo del tiempo. En el contexto del proyecciónlogístico y

comercial, cada métrica aporta una perspectiva diferente sobre la utilidad real del modelo en la toma de decisiones.

En este análisis, las métricas de evaluación se aplican exclusivamente durante la etapa de **validación final del modelo**, es decir, antes de generar las predicciones futuras. Esto permite comparar el desempeño de distintos algoritmos con datos históricos que no fueron utilizados durante el entrenamiento ni en la etapa de ajuste de hiperparámetros.

El objetivo es garantizar que los modelos seleccionados tengan una capacidad predictiva comprobada y estable antes de proyectar escenarios futuros con variables exógenas estimadas.

3.3.1 Coeficiente de determinación (R^2)

Definición: Es una medida estadística que indica qué proporción de la variabilidad de la variable dependiente puede ser explicada por el modelo.

Interpretación:

- Mide qué tan bien el modelo explica la varianza de los valores reales.
- Su rango va de $-\infty$ a 1, donde 1 representa una predicción perfecta.
- Un valor negativo de R^2 indica que el modelo tiene un desempeño peor que simplemente predecir la media de los datos.

¿Por qué es útil?

- Permite evaluar qué porcentaje de la variación en el volumen de carga (TEUs) puede ser atribuido a las variables independientes y a la estructura del modelo.
- Es útil para comparar diferentes modelos aplicados sobre el mismo conjunto de datos.

3.3.2 Raíz del error cuadrático medio (RMSE)

Definición: Es la raíz cuadrada del promedio de los errores al cuadrado entre los valores reales y los predichos.

Interpretación:

- Penaliza los errores grandes más fuertemente que los pequeños, debido a que los errores se elevan al cuadrado.
- Se expresa en las mismas unidades que la variable objetivo (por ejemplo, TEUs).

¿Por qué es útil?

- Es fundamental en logística, donde errores grandes —como una mala estimación de un pico de volumen de contenedores— pueden generar altos costos operativos o presión sobre la infraestructura.
- Es más sensible que el MAE ante fallas críticas del modelo.

3.3.3 Error absoluto medio (MAE)

Definición: Es el promedio de las diferencias absolutas entre los valores observados y los predichos.

Interpretación:

- Promedia la diferencia absoluta entre los valores reales y los predichos.
- Es menos sensible a valores atípicos que el RMSE.

¿Por qué es útil?

- Proporciona una medida directa e intuitiva del rendimiento del modelo de pronóstico, expresada en unidades reales.

- Es especialmente relevante para el seguimiento general del error sin sobrevalorar desviaciones extremas.

3.3.4 Error porcentual absoluto medio (MAPE)

Definición: Representa el error medio expresado como porcentaje del valor real observado

Interpretación:

- Normaliza el error como porcentaje del valor real.
- Puede verse distorsionado cuando los valores reales son muy pequeños, debido al efecto del denominador.

¿Por qué es útil?

- Facilita la comparación entre diferentes puertos, períodos o escalas de operación.
- Se utiliza frecuentemente en contextos de política pública o gestión empresarial, donde el error relativo es más relevante que el valor absoluto.

3.3.5 Precisión direccional (DA)

Definición: Es la proporción de veces en que el modelo predice correctamente si la variable aumentará o disminuirá en el siguiente período.

Interpretación:

- Mide la proporción de veces que el modelo predice correctamente la dirección del cambio (si el siguiente valor sube o baja).

¿Por qué es útil?

- Es particularmente importante en la planificación logística estratégica, donde anticipar la dirección de la tendencia (aumento o disminución del flujo de contenedores) puede ser más valioso que predecir el valor exacto.
- Sirve para alertar a los tomadores de decisiones y facilitar acciones preventivas, como reasignar recursos ante un posible aumento de la demanda.

3.3.6 Fórmulas

$$RMSE = \sqrt{\frac{1}{N}\sum_{t=1}^{N}(y_t - \hat{y}_t)^2}$$

$$MAE = \frac{1}{N}\sum_{t=1}^{N}|y_t - \hat{y}_t|$$

$$MAPE = \frac{1}{N}\sum_{t=1}^{N}(|y_t - \hat{y}_t| / |y_t|)$$

$$DA = \frac{1}{N}\sum_{t=1}^{N} d_t, \quad d_t = \begin{cases} 1 & if\ (y_{t+1} - y_t)(\hat{y}_{t+1} - y_t) \geq 0 \\ 0 & otherwise \end{cases}$$

Ilustración 4: Fórmulas de las métricas utilizadas para la evaluación del modelo de pronóstico: RMSE, MAE, MAPE y Precisión Direccional (DA). Fuente: Adaptado de Jiang, F., Xie, G., & Wang, S. (2021). *Forecasting Port Container Throughput with Deep Learning Approach*. CSAE 2021, ACM.

3.3.7 Ejemplo de uso de métricas

Métrica	Valor
R^2	0.411
MAE (TEU)	14,970
RMSE (TEU)	19,586
MAPE (\%)	3.72\%
DA (Precisión Direccional)	53.8\%

Ilustración 5: Resumen de las métricas de evaluación correspondientes al modelo representado en la Ilustración 1. Fuente: Elaboración propia.

El modelo GRU de la Ilustración 3 muestra un ajuste moderado, con un R^2 de 0.411, indicando que captura parte relevante de la variabilidad. El MAE de 14,970 TEUs y el RMSE de 19,586 TEUs revelan errores significativos en algunos puntos, especialmente en picos de demanda.

Aunque el MAPE de 3.72 % refleja buena precisión relativa, la precisión direccional (53.8 %) indica dificultades para anticipar correctamente la dirección de los cambios, como se observa en los puntos de inflexión. Se recomienda ajustar hiperparámetros o explorar modelos más complejos para mejorar la captura de patrones no lineales.

3.4 La importancia de los variables

Con el objetivo de identificar los determinantes más relevantes del comportamiento de la variable dependiente, se aplicó la técnica de Permutation Importance sobre el conjunto de validación. Esta metodología consiste en permutar aleatoriamente cada variable explicativa y observar el incremento resultante en el error de predicción. De esta manera, se estima la contribución marginal de cada variable a la capacidad predictiva del modelo.

En el caso de modelos lineales y de árboles (Ridge, Random Forest), esta técnica se implementó directamente sobre las predicciones del conjunto de prueba. Para redes neuronales recurrentes (LSTM y GRU), se aplicó una versión adaptada que permuta individualmente cada canal de entrada en la secuencia temporal, cuantificando la diferencia en el error cuadrático medio sobre los datos de validación. Este procedimiento se alinea con el enfoque propuesto por Freeborough y Van Zyl (2022), quienes aplicaron *permutation importance* a modelos GRU y LSTM en el contexto del proyecciónfinanciero, demostrando su validez como herramienta de interpretación en arquitecturas profundas.

3.5 Metodología de Proyección

Para la generación de pronósticos de la variable objetiva en el período de 12 meses, se siguió un esquema de proyección en múltiples pasos con generación previa de variables exógenas. En primer lugar, se construyeron modelos univariados individuales de proyección (suavizado exponencial aditivo con estacionalidad de 12 meses) para cada variable independiente, utilizando exclusivamente información histórica disponible hasta marzo de 2025. Cuando el ajuste no era viable, se aplicó un método estacional ingenuo como alternativa.

Una vez obtenidas las proyecciones de variables exógenas, se reentrenó cada modelo con la totalidad de los datos históricos (entrenamiento completo) utilizando los mejores hiperparámetros identificados en la fase de validación cruzada. Finalmente, se ejecutaron predicciones paso a paso para los doce meses del horizonte de proyección, actualizando en cada iteración los lags de la variable dependiente, las variables exógenas proyectadas y la información temporal (mes). Este enfoque garantiza una simulación realista del entorno predictivo, coherente con las mejores prácticas identificadas en la literatura reciente sobre proyección de variables meteorológicas y demanda logística (Suradhaniwar et al., 2021).

4) Conclusión

El desarrollo de esta plataforma representa un paso significativo hacia la modernización del análisis logístico nacional, mediante el uso de herramientas avanzadas de visualización, correlación y modelado predictivo. No obstante, uno de los principales desafíos identificados durante el proyecto fue la **limitada disponibilidad de datos históricos con alta frecuencia y consistencia**, una barrera común en entornos logísticos de América Latina. Esta limitación restringe la capacidad de entrenar modelos robustos, disminuye la precisión de los pronósticos a largo plazo y dificulta la identificación de patrones estacionales o tendencias estructurales confiables.

Pese a estas restricciones, el proyecto sienta las bases para una expansión futura. Los próximos pasos incluyen:

- **Escalar la metodología** a otros sectores logísticos clave, como el transporte aéreo, la operación del Canal de Panamá y los flujos comerciales en las zonas económicas especiales.

- **Ampliar el repositorio de datos**, incorporando nuevas fuentes y estableciendo mecanismos de actualización periódica.

- **Desarrollar y refinar modelos predictivos**, orientados a brindar herramientas de apoyo a la toma de decisiones estratégicas en el sector público, privado y académico.

El fortalecimiento de estas capacidades contribuirá a mejorar la comprensión integral del sistema logístico nacional y a consolidar una cultura de decisiones basadas en evidencia en Panamá.

5) References Bibliográficas

- **Autoridad Marítima de Panamá.** (s. f.). *Panamá, puerta de las Américas y líder marítimo mundial*. AMP. Recuperado de https://www.amp.gob.pa/noticias/notas-de-prensa/panama-puerta-de-las-americas-y-lider-maritimo-mundial/

- **Awah, P. C., Nam, H., & Kim, S.** (2021). Short-term forecast of container throughput: New variables application for the Port of Douala. *Journal of Marine Science and Engineering, 9*(7), 720. https://doi.org/10.3390/jmse9070720

- **Bergmeir, C., & Benítez, J. M.** (2012). On the use of cross-validation for time series predictor evaluation. *Information Sciences, 191*, 192–213. https://doi.org/10.1016/j.ins.2011.12.028

- **Chen, X., & Huang, L.** (2020). Port throughput forecast model based on Adam-optimized GRU neural network. In *Proceedings of the 2020 4th International Conference on Computer Science and Artificial Intelligence* (pp. 129–134). ACM. https://doi.org/10.1145/3445815.3445823

- **Cuong, T. N., You, S.-S., Long, L. N. B., & Kim, H.-S.** (2022). Seaport resilience analysis and throughput forecast using a deep learning approach: A case study of Busan Port. *Sustainability, 14*(21), 13985. https://doi.org/10.3390/su142113985

- **Freeborough, W., & van Zyl, T.** (2022). Investigating explainability methods in recurrent neural network architectures for financial time series data. *Applied Sciences, 12*(3), 1427. https://doi.org/10.3390/app12031427

- **Lee, G.-C., & Bang, J.-Y.** (2024). Forecasting container throughput of Singapore port considering various exogenous

variables based on SARIMAX models. *Forecasting, 6*(3), 748–760. https://doi.org/10.3390/forecast6030038

- **Nguyen, T. P., & Cho, G. S.** (2024). Forecasting the Busan container volume using XGBoost approach based on machine learning model. *Journal of Internet of Things and Convergence, 10*(1), 39–45. https://journal.kci.go.kr/kiots/archive/articleView?artiId=ART003054124

- **Suradhaniwar, S., Kar, S., Durbha, S. S., & Jagarlapudi, A.** (2021). Time series forecasting of univariate agrometeorological data: A comparative performance evaluation via one-step and multi-step ahead forecasting strategies. *Sensors, 21*(7), 2430. https://doi.org/10.3390/s21072430

- **Yuan, A. E., & Shou, W.** (2024). A rigorous and versatile statistical test for correlations between stationary time series. *PLoS Biology, 22*(8), e3002758. https://doi.org/10.1371/journal.pbio.3002758

- **Yuan, L.** (2019). *Machine learning approach to forecasting empty container volumes* (Master's thesis). Blekinge Institute of Technology, Faculty of Computing, Karlskrona, Sweden.

Georgia Tech Panama
Logistics Innovation & Research Center

Un centro de innovación de SENACYT
Secretaría Nacional de Ciencia, Tecnología e Innovación

CONTÁCTANOS

(+507) 395-3030

georgiatechpanama@gatech.pa

gatechpanama