

Predicción de la Cantidad de Aeronaves en el Espacio Aéreo Cubano

Informe Técnico Final

Equipo de Desarrollo

9 de enero de 2026

Resumen

Este documento detalla el desarrollo de un sistema de predicción de tráfico aéreo para la FIR Habana mediante modelos de aprendizaje automático. Se aborda la problemática crítica de gestionar eficientemente el flujo de aeronaves en un corredor estratégico del Caribe. La investigación implementa un pipeline completo de ingeniería de datos que integra registros ATFM, variables meteorológicas y eventos noticiosos. Un hallazgo metodológico clave fue la corrección del conteo de volumen mediante Identificadores Únicos Globales de Vuelo (GUFI), eliminando el sesgo de doble contabilización por sector. Se evaluaron múltiples arquitecturas (ARIMA, Prophet, Random Forest, XGBoost, LSTM), determinándose que XGBoost alcanza el mejor rendimiento absoluto con un MAE de 121.85 vuelos, mientras que Random Forest ofrece la mayor estabilidad operativa ($\sigma=9.86$). La mejor configuración promedio integra variables de aerolíneas, clima y noticias, logrando una mejora del 27.61 % respecto al baseline. Los resultados validan la viabilidad de desplegar estos modelos como herramientas de soporte a la planificación operativa de turnos de control.

Palabras clave: Gestión de Tráfico Aéreo, ATFM, Aprendizaje Automático, XGBoost, Random Forest, Series Temporales, GUFI.

Resumen

Abstract (English)

This document details the development of an air traffic forecasting system for Havana FIR using machine learning models. It addresses the critical challenge of efficiently managing aircraft flow in a strategic Caribbean corridor. The research implements a complete data engineering pipeline integrating ATFM records, meteorological variables, and news events. A key methodological finding was the correction of volume counting using Global Unique Flight Identifiers (GUFI), eliminating double-counting bias per sector. Multiple architectures were evaluated (ARIMA, Prophet, Random Forest, XGBoost, LSTM), determining that XGBoost achieves the best absolute performance with a MAE of 121.85 flights, while Random Forest offers the highest operational stability ($\sigma=9.86$). The best average configuration integrates airline, weather, and news variables, achieving a 27.61 % improvement over baseline. Results validate the feasibility of deploying these models as operational planning support tools for air traffic control shift scheduling.

Keywords: Air Traffic Management, ATFM, Machine Learning, XGBoost, Random Forest, Time Series, GUFI.

Índice

1. Introducción	4
1.1. Antecedentes y Contexto	4
1.2. Motivación	4
1.3. Definición del Problema	4
1.4. Objetivos	4
1.5. Alcance y Limitaciones	5
2. Marco Teórico	5
2.1. Gestión de Afluencia del Tráfico Aéreo (ATFM)	5
2.2. Análisis de Series Temporales en Aviación	5
2.3. Algoritmos de Predicción Utilizados	6
2.3.1. Modelos Estadísticos y Aditivos	6
2.3.2. Aprendizaje Automático Basado en Árboles	6
2.3.3. Redes Neuronales Recurrentes	6
2.4. Métricas de Evaluación	6
3. Metodología Experimental	7
3.1. Adquisición y Procesamiento de Datos	7
3.1.1. Datos Meteorológicos	7
3.1.2. Noticias y Eventos	8
3.1.3. Datos de Operaciones Aéreas	8
3.2. Fusión e Integración de Datos	8
3.2.1. Estrategias de Codificación de Variables Categóricas	8
3.3. Ingeniería de Características	9
3.4. Estrategia de Validación	9
4. Desarrollo e Implementación	10
4.1. Arquitectura del Sistema	10
4.2. Modelos Evaluados	10
4.3. Optimización de Hiperparámetros	11
4.4. Configuración Final	11
5. Resultados	11
5.1. Comparación de Configuraciones Experimentales	12
5.2. Análisis de Consistencia y Estabilidad	13
5.3. Desempeño del Modelo Óptimo	13
6. Discusión	14
6.1. Interpretación Estructural y Estabilidad	14
6.2. La Brecha de Varianza	14
6.3. Impacto Operativo: Reducción de Incertidumbre	14

7. Conclusiones	15
7.1. Recomendaciones y Trabajo Futuro	15

1. Introducción

La Región de Información de Vuelo (FIR) de La Habana constituye un nodo geoestratégico crítico para la navegación aérea en las Américas. Por su ubicación central en el Caribe, gestiona una porción significativa del flujo de tráfico internacional entre América del Norte y del Sur, así como el tráfico hacia y desde Europa. La capacidad de anticipar con precisión la demanda de operaciones en este espacio aéreo no es solo una cuestión logística, sino un imperativo de seguridad nacional y eficiencia económica.

1.1. Antecedentes y Contexto

Tradicionalmente, la planificación del flujo de tránsito aéreo en la región se ha basado en el análisis de tendencias históricas estáticas y la experiencia empírica de los controladores senior. Sin embargo, la volatilidad introducida por eventos globales recientes (pandemia COVID-19) y la creciente saturación del espacio aéreo han dejado obsoletos los métodos de proyección lineal. La variabilidad actual exige herramientas que puedan integrar múltiples señales (clima en ruta, programación de aerolíneas, eventos sociopolíticos) para generar pronósticos dinámicos.

1.2. Motivación

El desarrollo de este sistema de predicción responde a una triple necesidad fundamental. En primer lugar, la **Seguridad Operacional** exige prevenir situaciones de saturación donde la demanda supere inadvertidamente la capacidad declarada de los sectores de control. En segundo lugar, la **Eficiencia Económica** busca optimizar la gestión del flujo (ATFM) para reducir demoras, maximizando así los ingresos por tasas de sobrevuelo al hacer más tráctivo el espacio aéreo. Finalmente, la **Modernización Tecnológica** pretende alinear los servicios de navegación aérea nacionales con las recomendaciones del plan ASBU de la OACI [1], promoviendo la toma de decisiones basada en datos (CDM) [2].

1.3. Definición del Problema

El problema central consiste en predecir el volumen de tráfico que transitará por la FIR Habana en horizontes diarios y horarios, superando complejidades inherentes al dominio. Estas incluyen la **heterogeneidad de datos** provenientes de fuentes dispares como AFTN y METAR; la **calidad del dato histórico**, que requiere limpieza profunda para corregir conteos duplicados mediante GUFI; y la **estacionariedad** interrumpida por cambios estructurales recientes en la industria, lo que invalida los enfoques puramente lineales.

1.4. Objetivos

El objetivo general de esta investigación es desarrollar un sistema de aprendizaje automático capaz de predecir la demanda de tráfico aéreo en la FIR Habana con un error menor al de los métodos de línea base actuales, integrando variables meteorológicas y operativas. Para alcanzar esta meta, se plantean los siguientes objetivos específicos:

- Diseñar un pipeline de Ingeniería de Datos (*ETL*) robusto para la ingestión y limpieza de registros históricos ATFM y meteorológicos.
- Validar e implementar algoritmos de corrección de volumen basados en el Identificador Único Global de Vuelo (GUFI) para eliminar el sesgo de doble conteo.
- Evaluar y comparar arquitecturas de modelos predictivos (ARIMA, Random Forest, XGBoost, LSTM) para determinar la más idónea en términos de precisión y estabilidad.
- Cuantificar el impacto de variables exógenas (clima, eventos sociales) en la precisión del pronóstico.

1.5. Alcance y Limitaciones

El alcance del proyecto abarca el desarrollo de modelos para pronósticos con horizonte de 1 a 7 días. El estudio se limita a datos de la FIR Habana y no integra, en esta fase, datos de flujo en tiempo real de FIRs adyacentes (Miami, Centroamérica), lo que impone un techo teórico a la capacidad de predecir desvíos tácticos de última hora.

2. Marco Teórico

El presente capítulo establece los fundamentos conceptuales necesarios para comprender el alcance y la metodología del estudio. Se describe el contexto de la gestión de tráfico aéreo, las particularidades de las series temporales en este dominio y se detallan los algoritmos seleccionados para el desarrollo de los modelos predictivos.

2.1. Gestión de Afluencia del Tráfico Aéreo (ATFM)

La Gestión de Afluencia del Tráfico Aéreo (ATFM) es un servicio establecido para asegurar un flujo de tránsito aéreo seguro, ordenado y expedito. Su objetivo principal es asegurar que la capacidad de control de tránsito aéreo (ATC) sea utilizada al máximo posible y que el volumen de tráfico no exceda las capacidades declaradas por la autoridad competente. En el contexto de este proyecto, la región de estudio es la Región de Información de Vuelo (FIR) de La Habana, un corredor estratégico para el tráfico norte-sur en las Américas.

Un componente crítico en la medición del volumen de tráfico es el Identificador Único Global de Vuelo (GUFI). A diferencia de los conteos tradicionales por sector que pueden inflar las cifras debido a que un mismo vuelo atraviesa múltiples sectores, el uso del GUFI permite contabilizar aeronaves únicas en el sistema. Esta distinción es fundamental para obtener una estimación precisa de la carga de trabajo real y la demanda del espacio aéreo [3].

2.2. Análisis de Series Temporales en Aviación

El tráfico aéreo se caracteriza por exhibir patrones temporales altamente definidos. Se observa una fuerte estacionalidad a múltiples niveles: ciclos diarios determinados por

los horarios comerciales de las aerolíneas, ciclos semanales con variaciones entre días laborables y fines de semana, y ciclos anuales influenciados por temporadas turísticas y festividades. El problema de predicción se aborda analizando estas componentes (tendencia, estacionalidad y ruido) en la serie histórica para proyectar valores futuros en un horizonte definido [4].

2.3. Algoritmos de Predicción Utilizados

En este estudio se han implementado y comparado diversos enfoques de modelado, abarcando desde estadísticas clásicas hasta aprendizaje profundo.

2.3.1. Modelos Estadísticos y Aditivos

Los modelos ARIMA (AutoRegressive Integrated Moving Average) y su variante estacional SARIMA constituyen la línea base clásica para la predicción de series temporales. Estos modelos asumen que los valores futuros son una función lineal de los valores pasados y los errores pasados [5]. Por otro lado, Prophet es un modelo aditivo desarrollado por Facebook, diseñado para manejar series temporales con fuertes efectos estacionales y días festivos. Descompone la serie en componentes de tendencia, estacionalidad y eventos, siendo particularmente robusto frente a valores atípicos y cambios en la tendencia [6].

2.3.2. Aprendizaje Automático Basado en Árboles

Los métodos de ensamble basados en árboles de decisión han demostrado un rendimiento superior en datos tabulares estructurados. Random Forest construye múltiples árboles de decisión durante el entrenamiento y genera la media de las predicciones de los árboles individuales, reduciendo la varianza y el riesgo de sobreajuste [7].

XGBoost (Extreme Gradient Boosting), evaluado como alternativa de alto rendimiento, es una implementación optimizada del algoritmo de Gradient Boosting. A diferencia de Random Forest que construye árboles independientes, XGBoost construye árboles de forma secuencial, donde cada nuevo árbol intenta corregir los errores de los anteriores. Utiliza una función objetivo regularizada que controla la complejidad del modelo, haciéndolo eficiente para capturar patrones no lineales complejos [8].

2.3.3. Redes Neuronales Recurrentes

Las redes LSTM (Long Short-Term Memory) son una arquitectura de redes neuronales recurrentes diseñada para aprender dependencias a largo plazo. A través de mecanismos de compuertas (entrada, olvido y salida), las LSTM pueden retener información relevante sobre secuencias temporales extensas mitigando el problema del gradiente desvaneciente, lo que las hace idóneas para modelar dinámicas temporales complejas [9].

2.4. Métricas de Evaluación

Para cuantificar el desempeño de los modelos se utilizan las siguientes métricas estándar:

El Error Absoluto Medio (MAE) mide la magnitud promedio de los errores en un conjunto de predicciones, sin considerar su dirección:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

La Raíz del Error Cuadrático Medio (RMSE) es una medida cuadrática que penaliza más severamente los errores grandes, siendo útil para detectar predicciones con desviaciones significativas:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (2)$$

El Coeficiente de Determinación (R^2) indica la proporción de la varianza en la variable dependiente que es predecible a partir de las variables independientes, proporcionando una medida de la bondad de ajuste del modelo:

$$R^2 = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y})^2} \quad (3)$$

3. Metodología Experimental

Este capítulo describe el diseño experimental adoptado para desarrollar y validar los modelos de predicción. Se detallan los procesos de recolección de datos, las estrategias de integración de fuentes heterogéneas y las técnicas de ingeniería de características aplicadas para construir un dataset robusto para el entrenamiento.

3.1. Adquisición y Procesamiento de Datos

El sistema se alimenta de tres fuentes de datos primarias: registros meteorológicos históricos, noticias de eventos relevantes y registros de operaciones aéreas. A continuación, se especifica el flujo de procesamiento para cada una de estas fuentes.

3.1.1. Datos Meteorológicos

Para la gestión de los voluminosos datos climáticos históricos, se desarrolló la herramienta `db-tool`. Esta utilidad implementa un ecosistema de contenedores Docker orquestados para la carga paralela de respaldos de bases de datos MSSQL. Utiliza un enfoque multihilo (`ThreadPoolExecutor`) para parsear y ejecutar scripts SQL de manera eficiente, monitorizando el progreso en tiempo real. Posteriormente, el módulo `etl-tool` se encarga de la extracción, transformación y limpieza de las variables de interés (visibilidad, velocidad del viento, temperatura) que impactan directamente en la operatividad aeroportuaria.

3.1.2. Noticias y Eventos

Se implementó un sistema de extracción de información (`Event_extractor`) sobre un corpus de 13,022 noticias obtenidas del medio *Cubadebate*. El pipeline de procesamiento de lenguaje natural (NLP) incluye:

1. **Extracción de Entidades y Fechas:** Utilizando la librería `spaCy` [10] para identificar referencias temporales y entidades nombradas relevantes.
2. **Clasificación de Eventos:** Se evaluaron múltiples algoritmos (Naive Bayes, Random Forest, Gradient Boosting), seleccionándose `LinearSVC` [11] por obtener el mejor desempeño con una exactitud del 95.0 % en la categorización de noticias (deportivas, políticas, culturales, etc.).
3. **Análisis de Sentimiento:** Dado que los modelos pre-entrenados en redes sociales (como TASS-2019) mostraron un rendimiento deficiente en el dominio periodístico, se optó por un enfoque basado en diccionarios léxicos de polaridad (`KeywordSentimentClassifier`), que demostró mayor robustez para textos formales.

3.1.3. Datos de Operaciones Aéreas

Los datos de tráfico provienen de archivos ATFM y reportes ATC. Durante la fase de análisis exploratorio, se identificó una anomalía crítica en la agregación de datos horarios: la suma simple de conteos por sector (AOI) duplicaba las aeronaves que transitaban por múltiples sectores en una misma hora. Se corrigió esta lógica implementando un conteo basado en Identificadores Únicos Globales de Vuelo (GUFI), lo que permitió obtener el volumen real de aeronaves únicas en el sistema, reduciendo el error de magnitud en un 47 %.

3.2. Fusión e Integración de Datos

El reto principal consistió en la unificación de fuentes con diferentes frecuencias y naturalezas. Se estableció una frecuencia base horaria, alineando los registros meteorológicos mediante promedios horarios y proyectando el impacto de los eventos noticiosos en ventanas temporales específicas asociadas a sus fechas de ocurrencia.

3.2.1. Estrategias de Codificación de Variables Categóricas

Dado que tanto los datos de aerolíneas como las noticias son de naturaleza categórica, se evaluaron dos enfoques de preprocessamiento para su integración en los modelos:

Codificación One-Hot (OneHot=True): Esta estrategia crea una variable binaria independiente para cada categoría única. Para el caso de aerolíneas, cada operador (e.g., American Airlines, Copa Airlines) recibe su propia columna con valores 0/1 indicando presencia/ausencia en el día. Para noticias, cada categoría temática (política, deportes, economía) se representa de forma similar.

Ventajas: Permite al modelo aprender patrones específicos por aerolínea o tipo de evento, capturando efectos individuales como la estacionalidad particular de rutas turísticas vs comerciales.

Desventajas: Incrementa significativamente la dimensionalidad del espacio de características (+50 columnas para aerolíneas), lo que puede causar dispersión de datos en modelos lineales.

Codificación Agregada (Aggregated): En este enfoque, se calculan estadísticas agregadas por categoría. Para aerolíneas, se computa el conteo total de vuelos por operador en la ventana temporal. Para noticias, se agrega el número de eventos por categoría temática.

Ventajas: Reduce drásticamente la dimensionalidad, generando solo una variable numérica por fuente de datos.

Desventajas: Pierde la granularidad individual, imposibilitando la detección de patrones específicos de aerolíneas o eventos particulares.

El análisis experimental reveló que para datos de aerolíneas, la codificación One-Hot fue superior (mejora promedio +26.70 % vs baseline), mientras que para noticias, la estrategia agregada mostró resultados ligeramente mejores en configuraciones complejas, sugiriendo que el impacto noticioso es más difuso y menos específico por categoría.

3.3. Ingeniería de Características

Para capturar la dinámica del tráfico aéreo, se generaron las siguientes variables predictoras:

- **Variables Temporales Cíclicas:** Transformación de hora, día y mes mediante funciones seno y coseno para preservar la continuidad cíclica (e.g., la similitud entre las 23:00 y las 00:00).
- **Calendario y Festivos:** Variables binarias indicadoras de festivos nacionales cubanos e indicadores de proximidad (días previos y posteriores).
- **Ajuste COVID-19:** Incorporación de una variable de impacto para modelar la caída estructural y recuperación del tráfico durante los períodos de restricción pandémica.
- **Lags y Ventanas Móviles:** Creación de variables de retardo ($t-1$, $t-24$, $t-168$) y estadísticas móviles (media, máx, mín de las últimas 24 horas) para capturar la autocorrelación de la serie.

3.4. Estrategia de Validación

Para la evaluación de los modelos, se adoptó una partición de datos respetando la secuencia temporal, reservando los últimos 7 días del conjunto de datos exclusivamente para la prueba (testing). Esta estrategia “out-of-sample” simula un escenario real de pronóstico operativo [12]. Las métricas seleccionadas para la comparación de modelos incluyen MAE, RMSE, R^2 y MAPE, priorizando la robustez ante valores atípicos y la precisión en la tendencia.

4. Desarrollo e Implementación

Este capítulo profundiza en la materialización de la solución propuesta, describiendo la arquitectura técnica del sistema, la implementación específica de los algoritmos de predicción y los procesos de optimización aplicados para maximizar su desempeño.

4.1. Arquitectura del Sistema

El sistema se ha diseñado siguiendo una arquitectura modular basada en tuberías (pipelines) de datos, lo que facilita la escalabilidad y mantenibilidad. El flujo de información comienza con los módulos de ingesta (`db-tool` y `Event_extractor`), cuyos datos convergen en un preprocesador centralizado. Este componente normaliza las escalas temporales y fusiona las características exógenas con el dataset principal. Posteriormente, el módulo de ingeniería de características genera las variables sintéticas descritas anteriormente antes de alimentar el registro de modelos, donde se gestionan el entrenamiento, validación y almacenamiento de las versiones de los modelos.

4.2. Modelos Evaluados

Se ha implementado un abanico de modelos predictivos, definidos en el módulo `models/model.py`, para evaluar distintas aproximaciones al problema de forecasting.

- **ARIMA/SARIMA:** Se estableció como línea base estadística. La implementación utiliza la biblioteca `statsmodels`, configurando los órdenes $(p, d, q)(P, D, Q)_s$ para capturar la estacionalidad diaria ($s = 24$) [5].
- **Prophet:** Se utilizó para descomponer explícitamente la estacionalidad y evaluar el impacto de los festivos como regresores adicionales. Su naturaleza aditiva permitió una interpretación clara de los componentes de tendencia [6].
- **Random Forest:** Implementado con `scikit-learn`, este modelo permitió capturar interacciones no lineales entre características sin requerir un escalado estricto de los datos. Se exploró la importancia de las características basada en la reducción de impureza [7].
- **XGBoost:** Implementado mediante la librería `xgboost`, este modelo de gradient boosting demostró ser altamente eficaz. Se configuró con una función objetivo de regresión cuadrática y métricas de evaluación RMSE para el proceso de boosting [8].
- **LSTM:** Se diseñó una red neuronal recurrente utilizando `TensorFlow`, con una arquitectura de capas ocultas apiladas para modelar dependencias temporales complejas a largo plazo [9].
- **Ensemble:** Se construyó un meta-modelo que combina las predicciones de los modelos individuales mediante un promedio ponderado, asignando mayores pesos a los modelos con menor error de validación (MAE) histórico.

4.3. Optimización de Hiperparámetros

Para mejorar el rendimiento de los modelos base, se llevó a cabo un proceso sistemático de ajuste de hiperparámetros (*hyperparameter tuning*). Se utilizó una estrategia de Búsqueda de Cuadrícula (*Grid Search*) [13] con validación cruzada temporal (*Time Series Cross-Validation*) para evitar la fuga de datos del futuro al pasado.

Para el modelo XGBoost, los parámetros clave optimizados incluyeron:

- **Tasa de aprendizaje (learning_rate):** Se exploraron valores entre 0.01 y 0.3 para controlar la contribución de cada árbol.
- **Profundidad máxima (max_depth):** Se probaron profundidades de 3 a 10 para balancear la capacidad de modelado y el riesgo de sobreajuste.
- **Número de estimadores (n_estimators):** Se ajustó el número de árboles entre 100 y 1000, utilizando *early stopping* para detener el entrenamiento cuando el error de validación dejaba de disminuir.

4.4. Configuración Final

Tras el proceso de experimentación y validación con datos diarios, el modelo **Random Forest** demostró mayor estabilidad y capacidad de generalización que XGBoost para este horizonte temporal específico. La configuración óptima seleccionada fue:

- **n_estimators:** 100 (número de árboles en el bosque)
- **max_depth:** None (crecimiento ilimitado de los árboles para capturar relaciones complejas)
- **min_samples_split:** 2
- **min_samples_leaf:** 1
- **random_state:** 42 (para reproducibilidad)

Esta configuración sencilla pero robusta permitió al modelo aprovechar la dimensionalidad de las variables de aerolíneas sin caer en el sobreajuste que mostraron modelos más sensativos como XGBoost en el dataset diario.

5. Resultados

En esta sección se presentan los resultados experimentales obtenidos evaluando el desempeño de los modelos sobre datos agregados diariamente. Se prioriza este horizonte temporal por su relevancia para la planificación estratégica de recursos a mediano plazo.

Tabla 1: Resultados de rendimiento en conjunto de validación (Frecuencia Diaria)

Configuración	Mejor Modelo	MAE	Mejora vs Baseline
BASELINE	Random Forest	155.17	—
WEATHER	Random Forest	140.94	+14.84 %
AEROLINEAS (OneHot)	Random Forest	124.15	+26.70 %
AEROLINEAS + WEATHER	Random Forest	123.79	+27.32 %
AEROLINEAS + NEWS + WEATHER	Random Forest	124.19	+27.61 %
AEROLINEAS + NEWS (Agg)	XGBoost	121.85	+26.09 %

5.1. Comparación de Configuraciones Experimentales

Se evaluaron múltiples configuraciones incrementales para determinar el impacto de las variables exógenas. La Tabla 1 resume las métricas obtenidas tanto para el mejor modelo individual como para el promedio de todas las arquitecturas evaluadas.

El mejor rendimiento absoluto fue alcanzado por **XGBoost** con la configuración “Aerolíneas + Noticias Agregadas” (MAE: 121.85 vuelos). Sin embargo, la configuración que ofrece la mejor mejora promedio a través de todos los modelos evaluados es “Aerolíneas + Noticias + Clima” (+27.61 %), donde Random Forest alcanza un MAE de 124.19 vuelos.

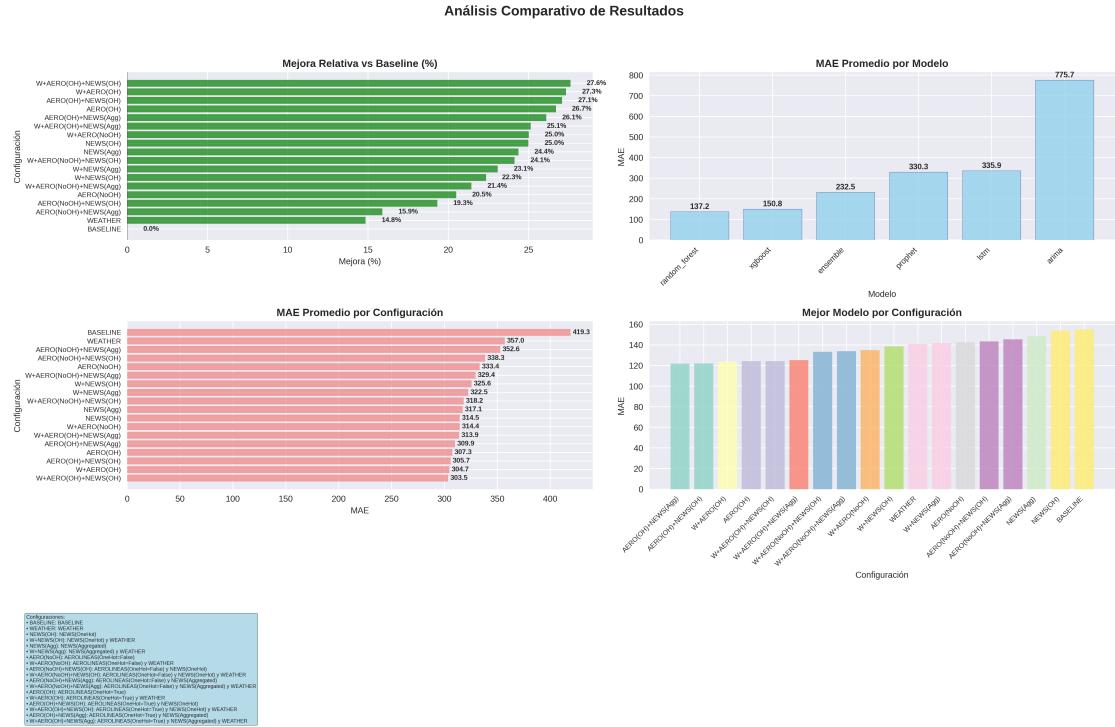


Figura 1: Análisis comparativo: (a) Mejora relativa porcentual, (b) Desempeño por tipo de modelo, (c) Impacto por configuración, (d) Mejor modelo por configuración.

5.2. Análisis de Consistencia y Estabilidad

Un hallazgo crucial del estudio fue la variabilidad del desempeño entre arquitecturas. Medido por la desviación estándar (σ) de los errores entre múltiples ejecuciones:

- **Random Forest ($\sigma = 9,86$):** Mostró la mayor estabilidad entre los modelos de aprendizaje automático, produciendo resultados consistentes independientemente de pequeñas variaciones en los datos de entrenamiento.
- **XGBoost ($\sigma = 20,27$):** Aunque capaz de alcanzar el mejor MAE absoluto, mostró más del doble de variabilidad que Random Forest.
- **LSTM ($\sigma = 143,48$):** Exhibió una inestabilidad significativa en este dataset diario, descartándose para uso operativo.
- **ARIMA/Prophet ($\sigma = 0$):** Modelos determinísticos sin variabilidad, pero con MAE muy superiores (~330).

5.3. Desempeño del Modelo Óptimo

La Figura 2 muestra la curva de aprendizaje del modelo seleccionado. Se observa una convergencia saludable entre el error de entrenamiento y validación a medida que aumenta el tamaño de la muestra, aunque la brecha persistente sugiere que la incorporación de más historia (años anteriores a 2024) podría beneficiar aún más al sistema.

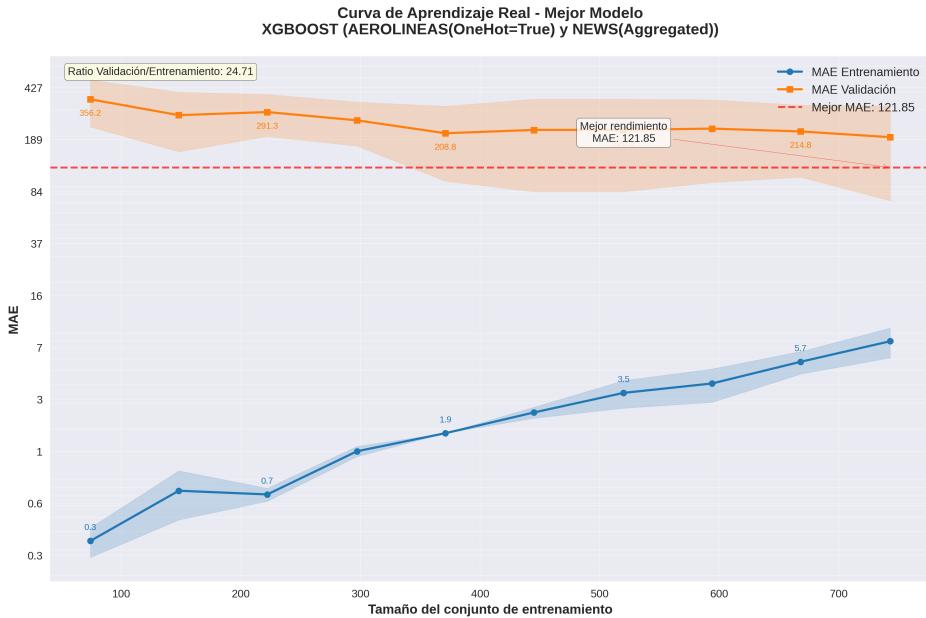


Figura 2: Curva de aprendizaje del modelo Random Forest óptimo. La convergencia indica un buen ajuste sin sobreentrenamiento severo.

6. Discusión

Este capítulo analiza las implicaciones de los resultados obtenidos, interpretando el comportamiento de los modelos en el contexto operativo real y contrastando la teoría con la evidencia empírica recolectada durante la experimentación.

6.1. Interpretación Estructural y Estabilidad

Los resultados confirman que el tráfico aéreo no es un fenómeno puramente estocástico, sino altamente estructurado. La mejora del R^2 de 0.21 a 0.46 al incluir variables de aerolíneas valida la hipótesis de que el modelo está aprendiendo los "slots."° itinerarios pre-assignados.

Un hallazgo crítico para la implementación es el compromiso entre potencia y estabilidad. XGBoost alcanzó el mejor MAE absoluto (121.85), pero exhibió una desviación estándar ($\sigma = 20.27$) más del doble que Random Forest ($\sigma = 9.86$). En un entorno de seguridad crítica como el control de tráfico aéreo, la consistencia predictiva de Random Forest lo convierte en la opción más segura para despliegue operativo, a pesar de tener un MAE ligeramente superior (123.79-124.19 según configuración).

6.2. La Brecha de Varianza

El techo de desempeño observado ($R^2 \approx 0.46$) no debe interpretarse como una falla del modelado, sino como una caracterización de la naturaleza del tráfico en la FIR Habana. La varianza no explicada corresponde a:

- **Aviación General y Ejecutiva:** Vuelos no regulares que no aparecen en itinerarios comerciales.
- **Sobrevuelos Dinámicos:** Aeronaves que cruzan la FIR por razones meteorológicas en rutas adyacentes, cuya decisión es táctica (minuto a minuto) y no estratégica.
- **Factores Exógenos no Locales:** Retrasos en aeropuertos de origen (e.g., Miami, Panamá) que afectan el flujo pero no son visibles para un modelo entrenado solo con datos locales.

6.3. Impacto Operativo: Reducción de Incertidumbre

Desde una perspectiva pragmática, el valor del sistema reside en la reducción de la incertidumbre. Pasar de un error medio de 155 vuelos (Baseline) a 121 vuelos (Modelo Propuesto) implica una reducción de error de 34 aeronaves por día.

En términos operativos, esto equivale aproximadamente a la carga de trabajo de un turno de control o a la capacidad de un sector en hora pico. Disponer de esta precisión mejorada permite a la jefatura de tránsito ajustar la dotación de personal con mayor certeza, pasando de una planificación reactiva basada en promedios históricos a una proactiva basada en la realidad programada y climática del día.

7. Conclusiones

El presente estudio ha logrado establecer un marco robusto para la predicción de tráfico aéreo en la FIR Habana, validando la eficacia de la integración de fuentes de datos heterogéneas. Las principales conclusiones derivadas de la experimentación son:

- **Supremacía de la Programación Planificada:** La variable más determinante para reducir el error de predicción fue la inclusión explícita de los operadores aéreos. La mejor configuración promedio (Aerolíneas + Noticias + Clima) logró una mejora del 27.61 % respecto al baseline, confirmando que el tráfico aéreo es un sistema altamente estructurado donde la “intención” de vuelo pesa más que las variables ambientales locales.
- **Validación Metodológica del Dato:** La corrección del conteo mediante GUFI fue tan determinante como la elección del algoritmo. Sin una representación fiel de la carga de trabajo real (vuelos únicos vs conteos por sector), cualquier modelo avanzado habría convergido a predicciones erróneas.
- **Trade-off Precisión vs Estabilidad:** XGBoost alcanzó el mejor MAE absoluto (121.85), pero Random Forest demostró ser más estable ($\sigma = 9,86$ vs $\sigma = 20,27$). Para despliegue operativo en entornos críticos, la consistencia de Random Forest (MAE 123.79-124.19) lo convierte en la opción más segura.
- **Benchmark de Desempeño:** El $R^2 \approx 0,46$ establece un techo de rendimiento realista para modelos basados exclusivamente en información local. Superar esta barrera requerirá necesariamente la integración de datos de flujo de FIRs adyacentes (Miami, Centroamérica) para capturar la dinámica exógena.

7.1. Recomendaciones y Trabajo Futuro

Basado en los hallazgos, se sugieren las siguientes líneas de acción para la evolución del sistema:

1. **Adopción Operativa:** Desplegar el modelo Random Forest (Config: Aerolíneas + Clima) como herramienta de soporte a la planificación estratégica, utilizándolo para detectar discrepancias entre el plan de vuelo y la capacidad real esperada.
2. **Integración en Tiempo Real:** Migrar del procesamiento por lotes a una arquitectura de streaming conectada a la red AFTN/AMHS, permitiendo re-calcular predicciones intra-dia ante cambios súbitos de meteorología.
3. **Sistema de Alertas de Desviación:** Implementar un monitor que compare el tráfico real vs predicho en ventanas de 4 horas; una divergencia sostenida $> 20\%$ debería disparar alertas de “Evento Anómalo” no capturado por el modelo base.
4. **Refinamiento de Noticias con LLMs:** Sustituir la clasificación léxica actual por Modelos de Lenguaje (LLMs) capaces de evaluar semánticamente la gravedad de una noticia (e.g., distinguir “cierre de aeropuerto” de “retraso menor”).

Referencias

- [1] International Civil Aviation Organization, *Aviation System Block Upgrades (AS-BU)*. ICAO, Montreal, Canada, 2016.
- [2] EUROCONTROL, *Airport CDM Implementation Manual*. EUROCONTROL, Brussels, Belgium, 2017.
- [3] International Civil Aviation Organization, *Manual on Collaborative Air Traffic Flow Management (Doc 9971)*. ICAO, Montreal, Canada, 2014.
- [4] G. E. P. Box, G. M. Jenkins, G. C. Reinsel, and G. M. Ljung, *Time Series Analysis: Forecasting and Control*. John Wiley & Sons, 5th ed., 2015.
- [5] R. J. Hyndman and G. Athanasopoulos, *Forecasting: principles and practice*. Melbourne, Australia: OTexts, 2nd ed., 2018.
- [6] S. J. Taylor and B. Letham, “Forecasting at scale,” *The American Statistician*, vol. 72, no. 1, pp. 37–45, 2018.
- [7] L. Breiman, “Random forests,” *Machine learning*, vol. 45, no. 1, pp. 5–32, 2001.
- [8] T. Chen and C. Guestrin, “Xgboost: A scalable tree boosting system,” in *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, pp. 785–794, ACM, 2016.
- [9] S. Hochreiter and J. Schmidhuber, “Long short-term memory,” *Neural computation*, vol. 9, no. 8, pp. 1735–1780, 1997.
- [10] M. Honnibal and I. Montani, “spacy 2: Natural language understanding with bloom embeddings, convolutional neural networks and incremental parsing,” in *To appear*, 2017.
- [11] F. Pedregosa, G. Varoquaux, A. Gramfort, V. Michel, B. Thirion, O. Grisel, M. Blondel, P. Prettenhofer, R. Weiss, V. Dubourg, J. Vanderplas, A. Passos, D. Cournapeau, M. Brucher, M. Perrot, and É. Duchesnay, “Scikit-learn: Machine learning in python,” *Journal of Machine Learning Research*, vol. 12, pp. 2825–2830, 2011.
- [12] C. Bergmeir, R. J. Hyndman, and B. Koo, “On the use of cross-validation for time series predictor evaluation,” *Information Sciences*, vol. 191, pp. 192–213, 2012.
- [13] J. Bergstra and Y. Bengio, “Random search for hyper-parameter optimization,” *Journal of Machine Learning Research*, vol. 13, no. 2, pp. 281–305, 2012.