

Conclusión Final del Proyecto

En este proyecto, se compararon cuatro enfoques distintos de series temporales (ARIMA, SARIMA, SimpleRNN y LSTM) para predecir el tiempo por vuelta de los pilotos, utilizando la carrera de Abu Dhabi 2021 como caso de estudio.

Resumen de Hallazgos:

1. Modelos Estadísticos (ARIMA/SARIMA):

- **ARIMA** demostró una precisión casi perfecta (**MSE: 0.0011**) al modelar una sola tanda limpia, probando ser ideal para capturar tendencias lineales simples.
- **SARIMA** fue capaz de manejar los ciclos de paradas en boxes ("estacionalidad") de la carrera completa. Su predicción final del "ritmo normal" fue lógicamente correcta, y el error (**MSE: 0.2769**) se debió a eventos externos impredecibles (el Safety Car), no a una falla del modelo.

2. Modelos de Deep Learning (RNN vs. LSTM):

- Ambos modelos se probaron con una arquitectura idéntica de 32 -> 16 -> 8 unidades.
- El modelo **LSTM** sufrió de **subajuste (underfitting)**, produciendo una predicción de línea recta. Esto sugiere que la alta complejidad interna del modelo LSTM fue contraproducente con un conjunto de datos de entrenamiento tan limitado (pocas vueltas), impidiendo que el optimizador encontrara un patrón útil.
- El modelo **SimpleRNN** fue el **claro ganador** del enfoque de Deep Learning. Al ser una arquitectura más simple, logró converger y **aprender con éxito el patrón** de degradación de los neumáticos, ajustándose notablemente bien a los datos reales (**MAE: 0.9717 segundos**).

Conclusión Clave:

Este proyecto demuestra el equilibrio fundamental entre la complejidad del modelo y el tamaño de los datos. Para este dataset, la SimpleRNN ofreció un mejor rendimiento que la LSTM al evitar los problemas de convergencia, mientras que SARIMA probó ser una herramienta robusta para entender los ciclos de carrera.

Próximos Pasos y Recomendaciones

Basado en este análisis, se proponen las siguientes acciones para mejorar la capacidad predictiva del equipo de estrategia:

- **1. Implementar la RNN como Línea de Base:** El modelo SimpleRNN ha demostrado ser una herramienta rápida y fiable para predecir el ritmo base de un piloto en condiciones normales. Se recomienda implementarlo como un indicador de referencia para la degradación esperada.

- **2. Optimizar el Modelo Ganador (RNN):** Dado que la SimpleRNN fue la arquitectura de DL más prometedora, el siguiente paso lógico es realizar una **optimización de hiperparámetros** (usando herramientas como Keras Tuner) para encontrar la configuración óptima (ej. número de neuronas, tasa de aprendizaje) y maximizar su precisión.
- **3. Enriquecer los Modelos con Datos Externos (ARIMAX/LSTM con Múltiples Entradas):** La principal debilidad de todos los modelos fue su incapacidad para predecir eventos externos (el Safety Car). Para una predicción de nivel de carrera, es crucial evolucionar a modelos **SARIMAX** o **LSTM multivariados** que incluyan datos exógenos como:
 - Estado de la Pista (Bandera Verde/Amarilla/SC).
 - Tipo y antigüedad del compuesto de neumático.
 - Distancia con el coche de delante (tráfico).
- **4. Incrementar el Dataset para LSTM:** La falla del modelo LSTM se debió probablemente a la escasez de datos. Con un conjunto de datos más grande (múltiples carreras o temporadas), la capacidad de memoria a largo plazo de la LSTM podría superar a la RNN para modelar patrones de degradación más sutiles.