

Darek Mathias: Modelo de predicción de congestión en aeropuertos europeos.

Usando un dataset de movimientos aéreos en Europa entre el 2016-2022 (<https://www.kaggle.com/datasets/umerhaddii/european-flights-dataset>) se elabora un modelo de predicción de congestión aérea basada en los movimientos IFR (instrumental flying rate) que es el tipo de movimiento aéreo que causa más retrasos de forma natural dada la naturaleza de las condiciones en la que los pilotos tienen que volar:

- Bloque 1: Carga de datos.

Se ha usado un bucket en s3 para cargar el csv. Posteriormente

-Bloque 2: EDA

se ha realizado un EDA ligero para saber que tipo de datos podemos esperar junto a una gráfica de correlación de datos para poder saber si los datos tienen algo que ver entre sí.

-Bloque 3: Preparación de datos (Ingeniería de datos)

Se han lidiado con valores Null en los datos de operaciones IFR sustituyéndolos por la mediana de los datos del conjunto.

-Bloque 4: entrenamiento y evaluación de modelos IA:

modelos escogidos: Linear regression, random forest.

se ha procedido con su evaluación, comparación de métricas y gráfica de predicción.

Problemas encontrados: Pese a haber indicado el uso de todos los núcleos disponibles en la cpu (`rf_model = RandomForestRegressor(n_estimators=100, n_jobs=-1, random_state=42)`) el modelo random forest tarda demasiado en compilar.