

Optimización del Consumo de Combustible en Operaciones Aéreas: Un Modelo Predictivo para la Gestión del Peso del Equipaje

Carlos Beytia¹[201921048–3] and Nicolás Nazar¹[202030539–0]

Departamento de Electrónica, Universidad Técnica Federico Santa María, San Joaquín, Chile

Abstract. En respuesta al imperativo global de reducir las emisiones de dióxido de carbono, nuestro equipo está iniciando una fase de investigación para desarrollar una estrategia que optimice el consumo de combustible en operaciones aéreas mediante la predicción precisa del peso del equipaje.

Con el objetivo de elaborar un modelo predictivo eficaz, pretendemos emplear técnicas avanzadas de aprendizaje automático que facilitarán el cálculo del peso del equipaje basado en diversas variables relevantes. Este enfoque podría alterar potencialmente la cantidad de combustible cargado en cada vuelo, permitiendo una gestión más adecuada de las emisiones de CO₂.

La metodología que proponemos implica la recolección y procesamiento de datos internos y externos, que abarcan operaciones de vuelo, detalles de los pasajeros y condiciones climáticas, entre otros factores. Planeamos utilizar bibliotecas de Python como Pandas y Scikit-learn para la limpieza y análisis de datos, con la posibilidad de implementar estrategias de regresión como regresión lineal múltiple o Regresión de Vectores de Soporte.

Una vez implementado, el modelo podría contribuir a reducir las emisiones de CO₂ y facilitar una gestión de combustible más precisa, lo cual podría traducirse en beneficios económicos. Además, el sistema podría tener la capacidad de adaptarse a una amplia gama de operaciones de vuelo, afectando potencialmente la experiencia del cliente, posiblemente a través de tarifas ajustadas debido a costos operativos optimizados.

1 Introducción

Introducción

En la era contemporánea, la mitigación del cambio climático ha emergido como uno de los desafíos más críticos a nivel global. En este contexto, la industria de la aviación juega un papel crucial, ya que es un significativo contribuyente a las emisiones de dióxido de carbono (CO₂). Una estrategia prometedora para reducir estas emisiones implica la optimización del combustible en vuelos, lo que se puede lograr mediante la gestión eficiente del peso que se lleva a bordo. El presente estudio se enfoca en el desarrollo de un modelo de Machine Learning (ML) que predice el peso del equipaje en bodega de los aviones, permitiendo así una planificación más precisa del combustible necesario para cada vuelo.

La relevancia de este trabajo radica en su potencial para contribuir significativamente a la reducción de la huella de carbono en la aviación, un sector donde las opciones de mitigación son limitadas y de alta importancia. A través de la aplicación de técnicas avanzadas de ML, este estudio busca no solo aportar en el ámbito ambiental, sino también ofrecer una herramienta práctica para la optimización de recursos en la industria aérea. El objetivo principal es desarrollar un modelo predictivo que, con base en datos históricos de check-ins de una aerolínea, estime con precisión el peso del equipaje en bodega para vuelos futuros. Este enfoque no solo apunta a una gestión más eficiente del combustible, sino también a una operación más sostenible y económicamente viable para las aerolíneas.

La investigación se centra en un conjunto de datos proporcionados por la Aerolínea X, que incluye información detallada sobre el peso del equipaje, fechas de vuelo, códigos de aeropuertos, y categorías de viajeros frecuentes, entre otros. Mediante el procesamiento y análisis de estos datos, se busca desarrollar un modelo que pueda prever con exactitud el peso total del equipaje para cada vuelo, tomando en cuenta diversas variables y patrones encontrados en los datos históricos.

2 Estado Del Arte

2.1 Flight Delay Prediction Based on Aviation Big Data and Machine Learning[1]

En el ámbito de la aviación, la capacidad de predecir con precisión diversos aspectos operativos es esencial para optimizar recursos, mejorar la eficiencia y garantizar la satisfacción del cliente. Con el auge de la era digital, la industria ha experimentado una inundación de datos provenientes de diversas fuentes, desde sistemas de vigilancia de tráfico aéreo hasta registros de pasajeros. Esta abundancia de datos ha abierto nuevas oportunidades para aplicar técnicas avanzadas de aprendizaje automático y análisis de datos en la solución de problemas complejos.

Un trabajo que destaca en este contexto es el realizado por Gui et al., titulado "Flight Delay Prediction Based on Aviation Big Data and Machine Learning".

En este estudio, el equipo abordó el desafío de predecir retrasos en vuelos, un problema que tiene implicaciones significativas tanto para las aerolíneas como para los pasajeros. Utilizando un enfoque basado en bosque aleatorio, lograron una precisión del 90.2% en la clasificación binaria de retrasos, un logro notable que demuestra el potencial del aprendizaje automático en este campo.

Lo que hace que este estudio sea particularmente relevante para nuestro trabajo es el conjunto detallado de atributos que se utilizaron para entrenar su modelo. Estos atributos abarcan desde condiciones meteorológicas, como la dirección y potencia del viento, hasta detalles operativos, como el aeropuerto de salida y llegada y los horarios programados de los vuelos. Además, se incorporaron datos de tráfico aéreo y otros factores contextuales que podrían influir en los retrasos.

Para nuestro objetivo de predecir el peso del equipaje en un vuelo, varios de estos atributos podrían ser relevantes. Las condiciones meteorológicas, por ejemplo, podrían influir en la decisión de un pasajero de llevar equipaje adicional, como ropa de abrigo o equipo especializado. Del mismo modo, los horarios de vuelo y la información del aeropuerto podrían dar una idea de la duración del viaje y, por lo tanto, de la cantidad de equipaje que un pasajero podría necesitar. Sin embargo, es crucial que adaptemos y seleccionemos estos atributos con cuidado, teniendo en cuenta las características únicas de nuestro problema y las diferencias entre predecir retrasos y predecir el peso del equipaje.

2.2 Lost and Found: Predicting Airline Baggage At-risk of Being Mishandled[3]

La gestión eficiente del equipaje en la industria de la aviación es un desafío crítico que ha ganado prominencia con el aumento constante de pasajeros aéreos. En particular, el manejo inadecuado del equipaje, especialmente durante las conexiones de vuelo, ha sido una preocupación creciente. Según un estudio reciente presentado por Herbert van Leeuwen y su equipo en "Lost and Found: Predicting Airline Baggage At-risk of Being Mishandled", se destaca que el número de equipajes mal manejados durante la transferencia a un vuelo de conexión es significativamente alto.

El trabajo de van Leeuwen y su equipo propone un modelo basado en Light Gradient Boosting Machine (Light-GBM) para identificar equipajes en riesgo de ser mal manejados. Este modelo alcanzó una precisión del 51.55% y un recall del 52.04% en el conjunto de prueba, con un F1 score del 51.79%. Es importante destacar que el modelo propuesto es de naturaleza binaria, clasificando los equipajes como "en riesgo" o "no en riesgo".

Lo que es particularmente relevante para nuestro estudio es el conjunto de atributos que se extrajeron para entrenar su modelo. Estos atributos se relacionan con dos aspectos principales: características a nivel de proceso y características a nivel de equipaje. Algunas de las características clave incluyen el mes y la hora del día, el número de pasajeros de transferencia, el número de equipajes de transferencia, y varios tiempos de procesamiento relacionados con la ruta del equipaje dentro del aeropuerto.

Para nuestro objetivo de predecir el peso del equipaje en un vuelo, es esencial considerar cómo estos atributos podrían influir en la cantidad de equipaje que los pasajeros deciden llevar. Por ejemplo, el tiempo de conexión programado y el tipo de conexión podrían dar una idea de la duración del viaje y, por lo tanto, de la cantidad de equipaje que un pasajero podría necesitar. Además, las características relacionadas con el proceso de transferencia de equipaje, como los tiempos de procesamiento y las rutas dentro del aeropuerto, podrían ofrecer información valiosa sobre las tendencias y patrones del equipaje.

El trabajo de van Leeuwen y su equipo proporciona una visión valiosa del estado actual del arte en la predicción y manejo del equipaje en la industria de la aviación. Su aplicación de técnicas avanzadas de aprendizaje automático, específicamente el Light-GBM, y la consecución de resultados precisos en la clasificación, nos brindan una base sólida sobre la cual podemos construir y adaptar para nuestro problema específico de predecir el peso del equipaje en vuelos.

2.3 A Model of Fuel Consumption Estimation and Abnormality Detection based on Airplane Flight Data Analysis[4]

La eficiencia en el consumo de combustible es un aspecto crítico en la aviación, no solo desde una perspectiva económica, sino también en términos de sostenibilidad y reducción de emisiones. En el estudio "A Model of Fuel Consumption Estimation and Abnormality Detection based on Airplane Flight Data Analysis" de Zhen Pan, Chengzhi Chi y Jingkai Zhang del Chinese Aeronautical Radio Electronics Research Institute, se aborda la estimación del consumo de combustible y la detección de anomalías en el sistema de combustible de los aviones.

Este trabajo destaca por su enfoque en el uso de datos de vuelo para entrenar una red neuronal BP (Back Propagation) con el objetivo de estimar el consumo de combustible. A diferencia de los modelos tradicionales basados en el principio de equilibrio energético, que requieren consultar gráficos de rendimiento, este estudio propone un modelo que se entrena directamente con datos de vuelo reales. Además, el modelo no solo estima el consumo de combustible para una fase del vuelo, sino que segmenta automáticamente el trayecto del vuelo en cinco rutas: despegue, ascenso, crucero, descenso y aterrizaje. Cada una de estas rutas se entrena en diferentes redes neuronales BP para estimar el consumo de combustible de cada fase.

El modelo considera múltiples factores que pueden influir en el consumo de combustible, incluyendo, pero no limitado a, la dirección y velocidad del viento, la inclinación, la aceleración longitudinal y transversal, entre otros. Estos factores, cuando se incorporan al modelo, permiten que las estimaciones de consumo de combustible se acerquen mucho más a los datos reales de consumo de combustible de vuelo.

Para nuestro propósito de predecir el peso del equipaje en un vuelo, es posible que podamos adaptar y aprovechar algunos de los atributos utilizados en este estudio. Factores como la dirección y velocidad del viento, la aceleración y la

inclinación podrían influir en decisiones logísticas relacionadas con el equipaje. Por ejemplo, condiciones meteorológicas adversas podrían llevar a restricciones de peso en el equipaje. Además, la metodología de segmentar un vuelo en diferentes fases y entrenar modelos específicos para cada fase podría ser una estrategia valiosa para abordar nuestro problema desde diferentes ángulos.

En resumen, el trabajo de Pan, Chi y Zhang ofrece una perspectiva innovadora sobre cómo abordar la estimación del consumo de combustible utilizando técnicas avanzadas de aprendizaje automático. Su enfoque basado en datos y la segmentación detallada del vuelo en diferentes fases proporciona una base sólida sobre la cual podríamos construir y adaptar para nuestro estudio sobre la predicción del peso del equipaje.

2.4 Temporal pattern attention for multivariate time series forecasting[5]

El artículo aborda el desafío de predecir series temporales multivariadas. Estas series temporales, como la predicción del consumo de electricidad, la producción de energía solar y las piezas de piano polifónicas, tienen numerosas aplicaciones valiosas. Sin embargo, las interdependencias complejas y no lineales entre los pasos temporales y las series complican esta tarea.

Para lograr una predicción precisa, es esencial modelar la dependencia a largo plazo en los datos de series temporales. Una solución propuesta en el artículo es el uso de redes neuronales recurrentes (RNNs) con un mecanismo de atención. Aunque el mecanismo de atención típico revisa la información en cada paso temporal anterior y selecciona información relevante para generar las salidas, tiene dificultades para capturar patrones temporales a lo largo de múltiples pasos temporales. Para abordar esto, los autores proponen usar un conjunto de filtros para extraer patrones temporales invariantes en el tiempo, similar a transformar los datos de series temporales en su "dominio de frecuencia". Luego, proponen un nuevo mecanismo de atención para seleccionar series temporales relevantes y usar su información de dominio de frecuencia para la predicción multivariada.

El estudio destaca por su enfoque innovador en la atención a patrones temporales, en lugar de simplemente centrarse en pasos temporales individuales. Además, introduce una red neuronal convolucional (CNN) para extraer información de patrones temporales de cada variable individual.

En cuanto a la relevancia para el problema de predecir el peso del equipaje en un vuelo, varios de los conceptos y técnicas presentados en este artículo podrían ser aplicables. Las condiciones meteorológicas, por ejemplo, son una serie temporal que podría influir en la decisión de un pasajero de llevar más o menos equipaje. La capacidad de modelar y predecir series temporales multivariadas con precisión podría ser invaluable para este propósito. Además, los atributos como la información del aeropuerto, los horarios de vuelo y las condiciones meteorológicas, que se mencionaron en el artículo anterior sobre retrasos en vuelos, podrían integrarse en un modelo basado en las técnicas presentadas en este artículo para predecir el peso del equipaje.

3 Metodología

La metodología de este estudio se centra en el desarrollo y aplicación de un modelo de Machine Learning para predecir el peso total del equipaje en bodega de los aviones. El proceso metodológico se divide en varias etapas esenciales:

3.1 Recolección y Preparación de Datos

La recolección de datos constituye la base de este estudio. Los datos fueron obtenidos de los registros de check-in en los counters de la Aerolínea que compartió los datos con el equipo con fines académicos de forma anónima, reflejando información precisa sobre los pasajeros y su equipaje. Los atributos recopilados incluyen:

- **scheduled_flight_departure_date**: Fecha de salida del vuelo.
- **origin_airport_iata_code** y **destination_airport_iata_code**: Códigos IATA de aeropuertos de origen y destino.
- **baggage_kg_1_qty** a **baggage_kg_5_qty**: Peso de hasta cinco piezas de equipaje por pasajero.
- **allowed_baggage_qty**: Cantidad de equipaje permitido por pasajero.
- **ffp_category**: Categoría en el programa de viajero frecuente.
- **month_sin**, **month_cos**, **day_sin**, **day_cos**, **year_diff**: Representación cíclica de la fecha.
- **destination_ANF**, **destination_CJC**, **destination_PMC**, etc.: Codificación one-hot de los 20 destinos principales.
- **feriado**: Indica si la fecha del vuelo coincide con un feriado en Chile.
- **ffp_BLACK**, **ffp_BLACK SIGNATURE**, **ffp_GOLD**, etc.: Codificación one-hot de las categorías FFP.

Se filtraron los datos para centrarse en vuelos con origen en Santiago (SCL) y se excluyeron registros con un peso de equipaje igual a 1 kg por posibles errores de registro. .

3.2 Procesamiento de Datos

Los datos pasaron por un proceso de limpieza y transformación. Se manejó la fecha de manera cíclica utilizando funciones seno y coseno para representar los meses y días, se aplicó one-hot encoding a las categorías de FFP y a los 20 destinos principales, y se añadió una columna indicando si la fecha del vuelo coincidía con un feriado en Chile.

3.3 Agrupación y Síntesis de Datos

Se agruparon los datos por vuelo, sumando el peso total del equipaje, el equipaje permitido y categorizando los pasajeros por su membresía en el programa de viajeros frecuentes. Esto redujo el tamaño del conjunto de datos de aproximadamente 11 millones de filas a alrededor de 20 mil filas, facilitando el manejo y análisis de los datos.

3.4 Selección del Modelo de Machine Learning

Se eligió un modelo basado en Redes Neuronales debido a su capacidad para modelar relaciones no lineales complejas entre las variables. Se utilizaron neuronas con función de activación ReLU, considerando que la variable objetivo (peso del equipaje) siempre es un valor positivo.

3.5 Entrenamiento y Optimización del Modelo:

Se empleó el algoritmo Adam para la optimización del modelo, dada su eficacia en diversos tipos de datos. La función de pérdida utilizada fue el Mean Squared Error (MSE), para penalizar de manera significativa los errores grandes, cruciales en la estimación del peso del equipaje.

3.6 Validación y Pruebas del Modelo

El modelo se validó utilizando un conjunto de datos representando el 10% del total. Se calculó el MSE y se analizó el número de predicciones con variaciones absolutas menores a 600 kg y aquellas con un error porcentual menor al 25% respecto a los valores reales.

4 Resultados

La implementación y evaluación del modelo de redes neuronales para predecir el peso del equipaje en bodega de aviones arrojó los siguientes resultados clave:

4.1 Desempeño del Modelo

El modelo mostró un Mean Squared Error (MSE) de 175318 en el conjunto de validación. Esto se traduce en que, en promedio, las predicciones del modelo se desviaron aproximadamente 418.47 kg (la raíz cuadrada de 175318) de los valores reales. Este nivel de precisión indica una capacidad razonable del modelo para estimar el peso del equipaje en bodega.

4.2 Análisis de Precisión de las Predicciones

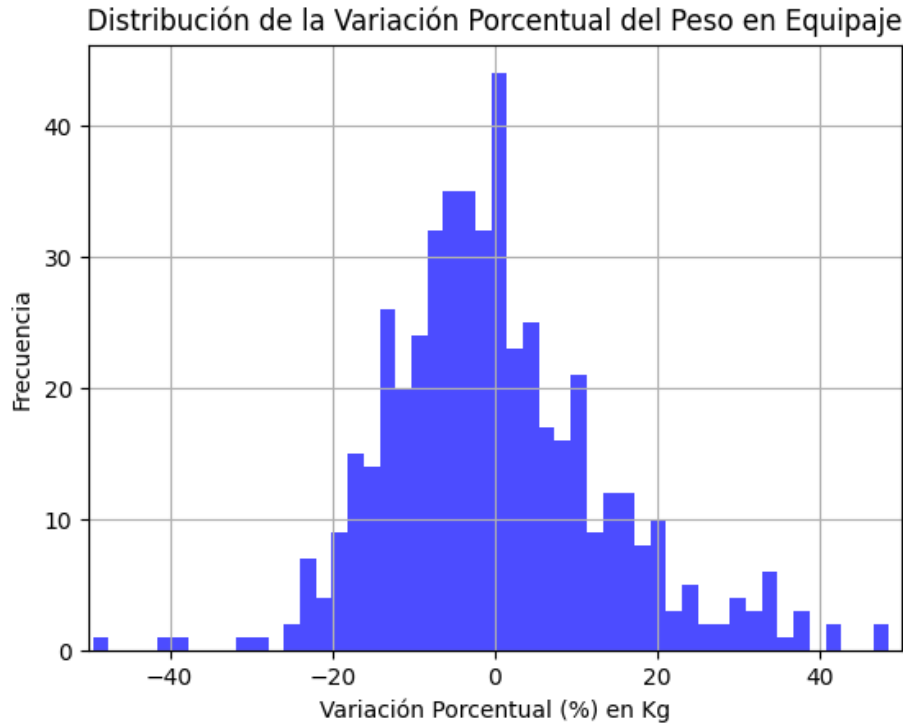


Fig. 1. Gráfico de la Distribución de la Variación porcentual del Peso en Equipaje

De un total de 500 predicciones evaluadas en el conjunto de prueba:

- 439 predicciones tuvieron una diferencia absoluta menor a 600 Kg respecto a los valores reales. Esto representa el 87.8% de las predicciones, una métrica relevante considerando que 600 Kg corresponde aproximadamente al 4% de la capacidad máxima de peso en bodega de un Boeing 787-8.[2]
- 451 predicciones presentaron un error porcentual menor al 25% en comparación con los valores reales, lo que equivale al 90.2% de las predicciones. Este nivel de precisión es significativo en el contexto de la planificación operativa de las aerolíneas [Fig1].

Estos resultados demuestran que el modelo tiene una capacidad notable para predecir con una precisión aceptable el peso del equipaje en bodega, lo que es esencial para la planificación eficiente del combustible y la operación de vuelos más sostenibles.

5 Discusión

Los resultados obtenidos en este estudio resaltan varias consideraciones y implicaciones importantes en el contexto de la predicción del peso del equipaje en bodega utilizando Machine Learning:

5.1 Interpretación del MSE y Precisión del Modelo

Aunque el MSE de 175318 sugiere una desviación significativa en términos absolutos, es esencial contextualizar este error en relación con las dimensiones operativas de la aviación comercial. La desviación promedio de 418.47 Kg, aunque notable, debe evaluarse en el marco del peso total de carga y capacidad de los aviones. En este contexto, el modelo muestra una precisión razonablemente buena, especialmente considerando la complejidad y variabilidad inherente a los datos de equipaje.

5.2 Comparación con Estudios Previos y Prácticas de la Industria

Esta investigación aporta al campo del Machine Learning aplicado a la aviación, un área donde la precisión en la predicción de pesos es crítica para la eficiencia del combustible y la reducción de emisiones de CO₂. La capacidad del modelo de alcanzar un 87.8% de predicciones con una variación menor a 600 Kg, y un 90.2% con un error menor al 25%, se destaca como un avance significativo en comparación con métodos convencionales, que a menudo dependen de estimaciones menos precisas basadas en promedios y experiencias pasadas.

5.3 Limitaciones del Estudio

Una limitación clave de este estudio es la dependencia de datos históricos específicos de una aerolínea, lo que podría afectar la generalización del modelo a otras aerolíneas o contextos geográficos. Además, la exclusión de registros con 1 Kg de peso en equipaje podría haber eliminado algunos datos válidos, aunque atípicos, afectando potencialmente la robustez del modelo.

5.4 Implicaciones Prácticas y Ambientales

La aplicación de este modelo en la planificación operativa puede conducir a una gestión más eficiente del combustible, lo que resulta en operaciones más rentables y sostenibles para las aerolíneas. La capacidad de predecir con mayor precisión el peso del equipaje puede traducirse en una carga de combustible más ajustada, reduciendo así las emisiones de CO₂ y contribuyendo a los esfuerzos globales de mitigación del cambio climático.

6 Conclusión

Este estudio representa un avance en la aplicación de técnicas de Machine Learning en la industria de la aviación, centrado específicamente en la predicción del peso del equipaje en bodega para optimizar el uso del combustible y contribuir a la reducción de emisiones de CO₂.

A través del desarrollo de un modelo basado en redes neuronales, se obtuvieron resultados favorables en la estimación del peso del equipaje, evidenciada por un Mean Squared Error de 175318 y una alta precisión en las predicciones, con un 87.8% de ellas dentro de una variación de 600 Kg y un 90.2% dentro de un margen de error porcentual del 25%.

Estos resultados no solo reflejan la capacidad del modelo para asistir en la planificación operativa eficiente de las aerolíneas, sino que también abren un camino prometedor hacia la realización de operaciones aéreas más sostenibles.

Sin embargo, es importante reconocer las limitaciones del estudio, incluyendo la dependencia de un conjunto de datos específico de una aerolínea y las posibles restricciones en la generalización de los resultados.

Asimismo, se identifica un potencial considerable para futuras investigaciones, donde la incorporación de variables adicionales, como condiciones climáticas y la exploración de modelos más avanzados de Machine Learning, podría mejorar sustancialmente la precisión y aplicabilidad del modelo.

Finalmente, este estudio no solo aporta a la eficiencia operativa y sostenibilidad en la aviación, sino que también establece una base sólida para futuras innovaciones en el uso de Machine Learning en la gestión y planificación de vuelos.

7 Trabajo futuro

El presente estudio abre varias avenidas para investigaciones futuras en el campo de la optimización de operaciones aéreas mediante el uso de Machine Learning. Las siguientes son algunas de las áreas prometedoras para futuros trabajos:

7.1 Incorporación de Variables Climáticas

Ampliar el conjunto de datos para incluir variables climáticas como temperatura, precipitaciones y velocidad del viento tanto en el aeropuerto de origen como en el de destino. Esto podría mejorar la precisión del modelo al tomar en cuenta factores ambientales que afectan el peso y la distribución del equipaje, así como el consumo de combustible.

7.2 Generalización del Modelo a Diferentes Aerolíneas y Contextos Geográficos

Realizar estudios adicionales utilizando datos de varias aerolíneas y de diferentes regiones geográficas para evaluar y mejorar la capacidad de generalización del modelo.

7.3 Integración con Sistemas de Gestión de Vuelos en Tiempo Real

Explorar la integración del modelo en sistemas de gestión de vuelos en tiempo real para proporcionar estimaciones dinámicas que puedan ajustarse a cambios de última hora en las operaciones de vuelo.

7.4 Desarrollo de Modelos Predictivos Más Avanzados

Investigar el uso de técnicas más avanzadas de Machine Learning, como redes neuronales profundas o aprendizaje automático reforzado, para mejorar aún más la precisión y eficacia del modelo.

7.5 Evaluación del Impacto Ambiental y Económico

Realizar estudios para evaluar el impacto real del modelo en términos de reducción de emisiones de CO₂ y eficiencia económica para las aerolíneas, proporcionando así una justificación más sólida para su implementación a gran escala.

7.6 Estudios de Impacto de la Implementación del Modelo

Investigar los cambios operativos y logísticos necesarios para implementar el modelo en la práctica, así como evaluar los posibles desafíos y resistencias en la industria de la aviación.

References

1. Guan Gui, Fan Liu, J.S.J.Y.Z.Z.D.Z.: Flight delay prediction based on aviation big data and machine learning. In: IEEE TRANSACTIONS ON VEHICULAR TECHNOLOGY, VOL. 69, NO. 1. pp. 140–150 (2020). <https://doi.org/doi:10.1109/TVT.2019.2954094>.
2. IAGCargo: BOEING 787-8 (2011), <https://www.iagcargo.com/es/fleet/boeing-787-8/>, Consultado el: 25 de noviembre, 2023]
3. Leeuwen, H., Zhang, Y., Zervanou, K., Mullick, S., Kaymak, U.: Lost and found: Predicting airline baggage at-risk of being mishandled. pp. 1–9 (01 2020). <https://doi.org/10.5220/0008977801720181>
4. Pan, Z., Chi, C., Zhang, J.: A model of fuel consumption estimation and abnormality detection based on airplane flight data analysis. In: 2018 IEEE/AIAA 37th Digital Avionics Systems Conference (DASC). pp. 1–6 (2018). <https://doi.org/10.1109/DASC.2018.8569796>
5. Shih, S.Y., Sun, F.K., Lee, H.y.: Temporal pattern attention for multivariate time series forecasting. Machine Learning **108**, 1421–1441 (2019). <https://doi.org/10.1007/s10994-019-05815-0>