**Aplicación de un modelo predictivo de consumo de combustible a partir de machine learning y random forest para una aerolínea comercial colombiana**

**Aplicação de um modelo preditivo de consumo de combustível usando aprendizado de máquina e random forest para uma companhia aérea comercial colombiana**

**Application of a predictive fuel consumption model using machine learning and random forest for a Colombian commercial airline**

**Resumen:**

En la industria aeronáutica, los costos de producción del hidrocarburo Jet A1 han experimentado un incremento en los últimos años, lo cual ha repercutido en un alza de los gastos para los operadores aéreos. Esta investigación se propone crear un modelo de predicción del consumo de combustible para vuelos regionales, basándose en los datos recolectados de vuelos de una aerolínea colombiana durante el período 2018-2019. Para la creación del modelo predictivo, se utilizó la biblioteca Sci-kit Learn del lenguaje de programación Python, y el enfoque del problema se realizó desde una perspectiva de 'problema inverso'. Luego se procedió con la ingeniería de características para mejorar la calidad del conjunto de datos obtenido y permitir una mayor precisión en la predicción. Se implementó el modelo de predicción para las variables identificadas como dependientes e independientes y, finalmente, se evaluó su rendimiento utilizando las métricas de error absoluto promedio (MAE), error cuadrático medio (MSE) y raíz del error cuadrático medio (RMSE). Los resultados indican que el modelo es eficiente para la predicción del consumo de combustible.

**Palabras clave:** Machine learning; Random Forest; pronóstico de combustible; métricas de evaluación; inteligencia artificial; feature engineering.

**Resumo**

Na indústria aeronáutica, os custos de produção do hidrocarboneto Jet A1 têm aumentado nos últimos anos, o que tem levado a um aumento nos gastos para os operadores aéreos. Esta pesquisa tem como objetivo criar um modelo de previsão do consumo de combustível para voos regionais, com base em dados coletados de voos de uma companhia aérea colombiana durante o período de 2018-2019. Para a criação do modelo preditivo, foi utilizada a biblioteca Sci-kit Learn da linguagem de programação Python, e a abordagem do problema foi realizada a partir de uma perspectiva de "problema inverso". Em seguida, foi realizada a engenharia de características para melhorar a qualidade do conjunto de dados obtido e permitir uma maior precisão na previsão. O modelo de previsão foi implementado para as variáveis identificadas como dependentes e independentes e, finalmente, seu desempenho foi avaliado usando métricas de erro médio absoluto (MAE), erro quadrático médio (MSE) e raiz do erro quadrático médio (RMSE). Os resultados indicam que o modelo é eficiente na previsão do consumo de combustível.

**Palavras-chave**: Machine learning; Random Forest; previsão de combustível; métricas de avaliação; inteligência artificial; feature engineering.

**Abstract**

In the aeronautical industry, the production costs of Jet A1 hydrocarbon have experienced an increase in recent years, resulting in higher expenses for airline operators. This research aims to create a predictive model for fuel consumption on regional flights based on data collected from flights of a Colombian airline during the period 2018-2019. The predictive model was developed using the Sci-kit Learn library in the Python programming language, approaching the problem from an inverse problem perspective. Feature engineering was performed to enhance the quality of the obtained dataset and allow for improved prediction accuracy. The predictive model was implemented for both dependent and independent variables, and its performance was evaluated using metrics such as Mean Absolute Error (MAE), Mean Squared Error (MSE), and Root Mean Squared Error (RMSE). The results indicate that the model is efficient in predicting fuel consumption.

**Key Words:** Machine learning; Random Forest; fuel forecasting; evaluation metrics; artificial intelligence; feature engineering.

**Introducción**

La demanda de pasajeros ha tenido un crecimiento aproximado de 5% año tras año desde 2011 a 2019 en el territorio nacional de la República de Colombia (Girón L., 2023). Ahora bien, la pandemia del COVID-19 tuvo gran repercución a nivel global lo que presentó una caída porcentual en demanda de pasajeros del 73%, sin embargo a partir de la reactivación de la economía la demanda de pasajeros ha tenido un crecimiento anual de alrededor del 4% (Viñas, L. C., & Gómez, M., 2023). Este crecimiento es importante analizarlo en los planes o estrategias a futuro de los operadores aeronáuticos colombianos y mundiales, con el proposito de pronosticar los costos de operación y la rentabilidad a periodos de corto, mediano y largo plazo.

Los aviones a reacción modernos utilizados en la aviación comercial mundial utilizan el hidrocarburo del queroseno, también llamado JET A-1 (Orellano Lasprilla, J. L., 2023). Este combustible ha presentado un aumento considerable, cerca del 250% en periodos de 4 años a nivel mundial (Morrison, Bonnefoy, Hansman, & Sgouridis, 2011), a causa de esta problemática que representa un porcentaje considerable al costo final por tiquete al usuario final, los operadores aeronáuticos deben velar el uso óptimo del producto químico, y evitar generar vuelos vacios que representen perdidas economicas.

La optimización del consumo de combustible es categorizada como un *inverse problem*, según Gallet et al (2022) el concepto se refiere a la fórmula que relaciona el sistema y los datos de un ambiente especifico. En el caso del pronostico de consumo de combustible, se ha esquematizado el montaje el cual de puede evidenciar en la Figura 1, donde existen tres escenarios, el sistema, el cual representa el consumo del carburante, los datos que son las variables obtenidas de la base de datos, y el método de pasar entre estos dos, el cual se denominó como *inverse problem* o la IA desarrollada.

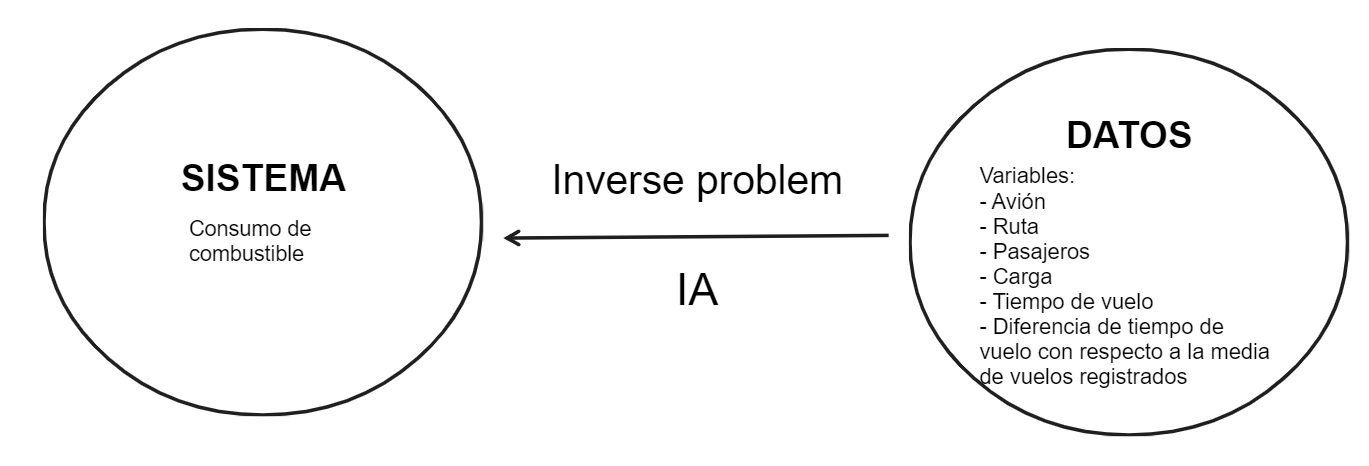


Figura 1. Esquema de inverse problem

Gracias a la idealización del concepto del *inverse problem*, muchas de las teorias físicas que rigen nuestro día a día existen, el caso de la representación de la fórmula de la atracción gravitacional entre dos masas existe a razón de que se tenia un sistema, la tierra y su satélite natura, y los datos como la masa y distancia, logrando explicar este comportamiento por una fórmula y así explicar la atracción gravitacional de los cuerpos celestes (Pizlo, Z., 2001). La IA generada es la fórmula que explica la relación entre los datos y el consumo de combustible de la aeronave objeto de estudio de la aerolinea.

El manejo de los datos independientes como el avión, ruta, numero de pasajeros, cantidad de carga, tiempo de vuelo, diferencia del tiempo de vuelo en relación a la media fue realizado a traves del concepto de ingeniería de características, conocido en inglés como *feaure engineering*. La ingeniería de caracteristicas es una serie de caracteristicas que enfoca sus esfuerzos en mejorar la calidad de los datos que se usen como *dataset* en un modelo de *machine learning*, estas técnicas conceptualizan desde el ingreso de datos faltantes a la base de datos hasta el escalamiento de los datos con el objetivo de ajustarlos a una misma escala y exista una mejor correlación (M. Panda, A. A. A. Mousa & A. E. Hassanien, 2021).

La inteligencia artificial es un campo integrado en las ciencias computacionales, su finalidad es plantear modelos de predicción con un esquema paralelo a las neuronas biologicas del ser humano, mediante una serie de redes neuronales, donde ingresa información, tiene sus propios pesos y genera un resultado, en la Figura 2 se puede observar la esquematización de una red neuronal con la analogia conceptual cientifica a una neurona biologica del ser humano. La IA permite generación de modelos de predicción y autoaprendizaje en diferentes escenarios, desde la visión computarizada hasta el reconocimiento de voz, existen diferentes tipos de aprendizaje dentro de la IA, cada uno tiene las ventajas y desventajas permitiendo una variedad de modelos IA para cada escenario (Márquez Díaz, J., 2020).

Diagrama

Descripción generada automáticamente

Figura 2. Esquematización de red neuronal

Dentro del estudio de IA, el aprendizaje automático, o en inglés *Machine Learning*, entra en acción, es un proceso que se ha utilizado para generar el modelo de predicción, a razón de que su principal funcion es el aprendizaje de las computadoras mediante el entrenamiento de algoritmos con datos especificados en su epoca de entrenamiento, con el fin de realizar tareas de clasificación, generación o conceptualización de nuevos datos (Janiesch, C., Zschech, P. & Heinrich, K., 2021). La selección del aprendizaje automático genera una gran ventaja cuando se deben realizar tareas repetitivas donde sus datos tengan relacióna al resultado esperado.

Profundizando en los diferentes secciones de *Machine Learning*, el aprendizaje supervisado o en inglés *supervised learning* es un modelo de predicción entrenado con datos etiquetados con anterioridad, con el objetivo de generar un modelo de IA capaz de clasificar o generar datos con etiquetas usadas en una escala física; el algoritmo es usado para análisis de texto (NLP), visión computacional o generación de datos a partir de la relación de los mismos (Baevski, A et al., 2022). A razón de la capacidad de generación de datos del aprendizaje supervisado, fue posible generar el pronostico de consumo de combustible mediante un algoritmo especializado para los datos que se habian obtenido.

La idealización de un modelo de predicción implica de igual manera seleccionar la clasificación correcta de aprendizaje automatico supervisado. Random Forest fue el algoritmo adecuado para la generación de los datos dependientes, o la cantidad de combustible para efectos prácticos. El algoritmo tiene la misión de generar árboles de desición sobre un conjunto de datos de entrenamiento, los cuales generan trabajan con un subconjunto aleatorio de variables y cada árbol llega hasta su máxima extensión; luego de obtener un resultado por cada árbol, se realiza un proceso de *soft voting* si la variable es categorica o de promedio si es númerica, permitiendo una precisión elevada a comparación de otros modelos de predicción similares (Espinosa-Zúñiga, J., 2020). En la Figura 3 se puede visualizar de manera gráfica el proceso de un algoritmo de Randon Forest para la generación o clasificación de datos.

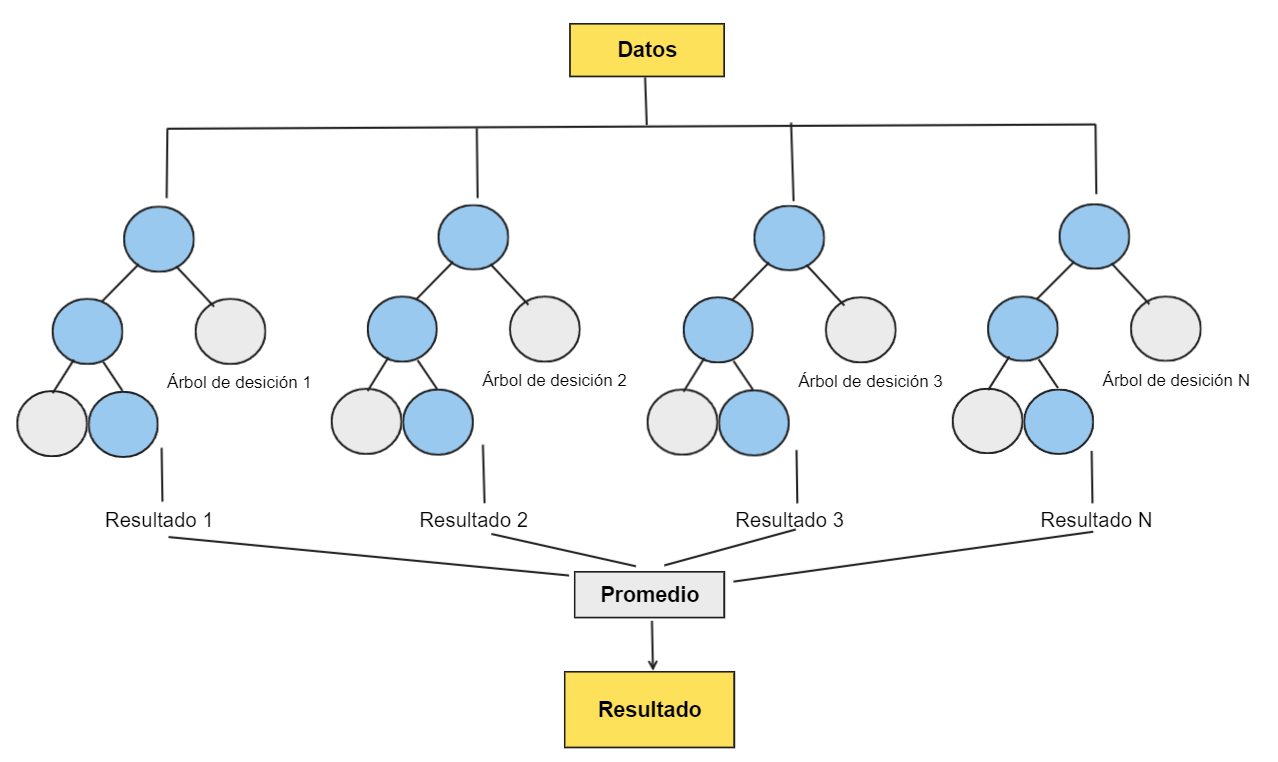


Figura 3. Esquema de Random Forest

En la Figura 4 se observa las secciones de los diferentes algoritmos que componen el objetivo del presente artículo.

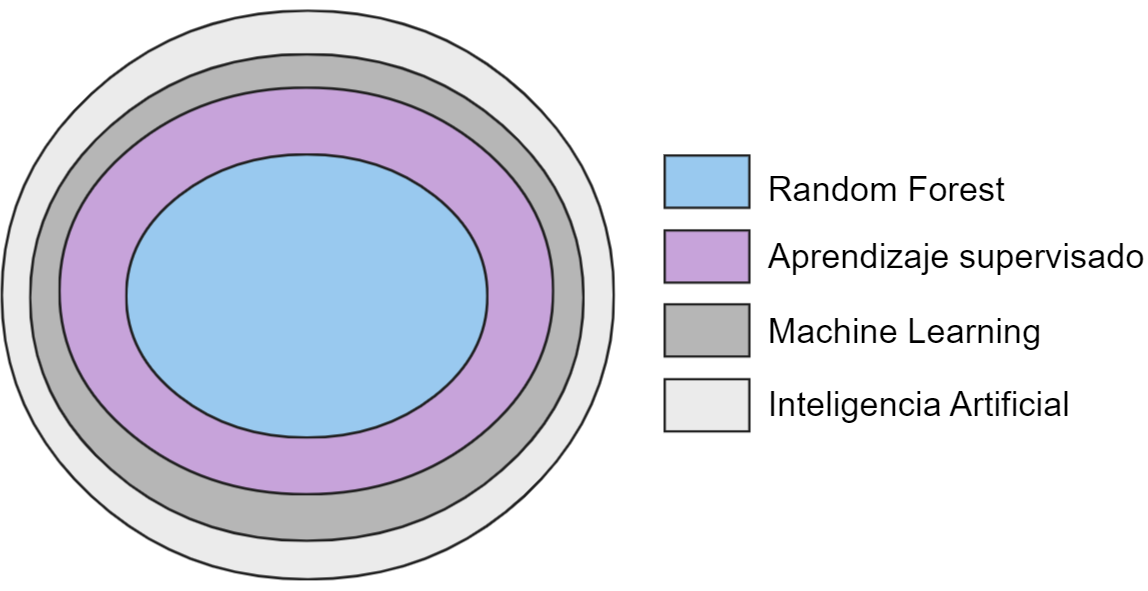


Figura 4. Esquema de IA a Random Forest

Finalmente la evaluación de un modelo de predicción se realiza mendiante métricas estadisticas, las más conocidas son MAE, MSE y RMSE. Las tres métricas son la evolución de leyes de lógica que a su vez provienen de las leyes de la lógica, MAE toma la diferencia entre el resultado esperado y el resultado obtenido, realiza una sumatoria y lo divide entre el numero total de cantidad de la muestra, el resultado se presenta en las unidades utilizadas y su ventaja radica en el fácil análsis; el MSE y RMSE son similares, la primera corresponde a la diferencia cuadrática de los valores esperado y obtenidos, haciendo sumatoria de estas sustracciones y la división con el total de unidades de la muestra, mientras que la segunda toma el valor de MSE y realiza el proceso de radicación, las ventajas de utilizar estas dos métricas radica en la penalización de los errores más grandes, sin embaro el MSE no utiliza las unidades esperadas, es razón primordial que el RMSE sea más utilizado que el anterior (Hodson, T. O, 2022). En todas las métricas expuestas, se busca que el resultado sea lo más cercano a 0, lo que indicaría que la diferencia entre los resultados esperados y obtenidos sea nula, indicando que el modelo de predicción es perfecto. En la Tabla 1, se observa la comparación entre las tres métricas utilizadas.

Tabla 1. Métricas de modelo de predicción

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Métrica** | **Fórmula** | **Unidades** | **Ventaja** | **Desventaja** |
| MAE |  | Misma de los datos. | Mejor comprensión, uso comercial. | No penaliza errores grandes. |
| MSE |  | Misma de los datos al cuadrado. | Penalización de errores grandes. | No maneja las mismas unidades que los datos. |
| RMSE |  | Misma de los datos. | Penalización de errores grandes, unidades iguales que los datos. | Posible sesgo a errores grandes poco frecuentes. |

**Método**

La metodología propuesta en este estudio se fundamenta en el enfoque CRISP-MD, estructurado en cuatro etapas principales, tal como se ilustra en la Figura 5. Este enfoque sistemático comienza con un análisis exploratorio de datos, donde se examina detenidamente el conjunto de datos para comprender sus características y patrones subyacentes. A continuación, se procede a la identificación de las características clave de los datos, lo que permite una mejor comprensión de las variables significativas que contribuyen al fenómeno en estudio. La tercera etapa consiste en la creación del modelo, donde se emplean técnicas de aprendizaje automático o estadísticas para desarrollar una representación matemática del fenómeno. Finalmente, se realiza una evaluación exhaustiva del modelo, aplicando métricas y pruebas relevantes para determinar su precisión, robustez y aplicabilidad en contextos específicos. Esta metodología integral garantiza un enfoque sistemático y riguroso para el análisis de datos, alineado con los estándares científicos actuales (Schröer, C., Kruse, F., & Gómez, J. M., 2021).

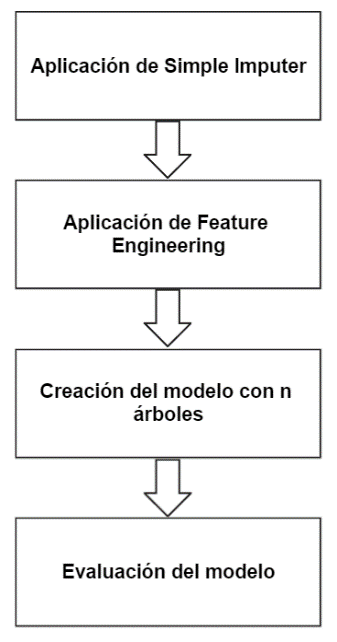


Figura 5. Flujograma de creación del modelo

**Aplicación Simple Imputer:**

Durante la fase de análisis exploratorio de datos (AED), se identificó una deficiencia en el conjunto de datos, particularmente en relación con la falta de información en ciertas variables críticas. Esta carencia requirió la implementación del primer procedimiento de ingeniería de características, utilizando una clase específica del módulo sci-kit learn en Python. Dicha clase ofrece varias estrategias para imputar los valores ausentes, como la media, la mediana, la moda o una constante predefinida. Tras un cuidadoso examen y evaluación de estas metodologías, se determinó que la imputación mediante la media era la más congruente con la naturaleza de los datos en cuestión. Por lo tanto, esta estrategia se aplicó a todas las columnas relevantes en las que se había detectado la omisión de información, con el objetivo de prevenir cualquier posible sesgo en futuras interpretaciones o análisis, en caso de que el modelo asumiera de manera predeterminada un valor nulo o cero para los datos faltantes.

El *feature* implementado se puede evidenciar en la Figura 6, la cual presenta un fragmento del código en Python manejado el *dataframe*, variable utilizada en el lenguaje de programación mencionado para base de datos en *Machine Learning*.

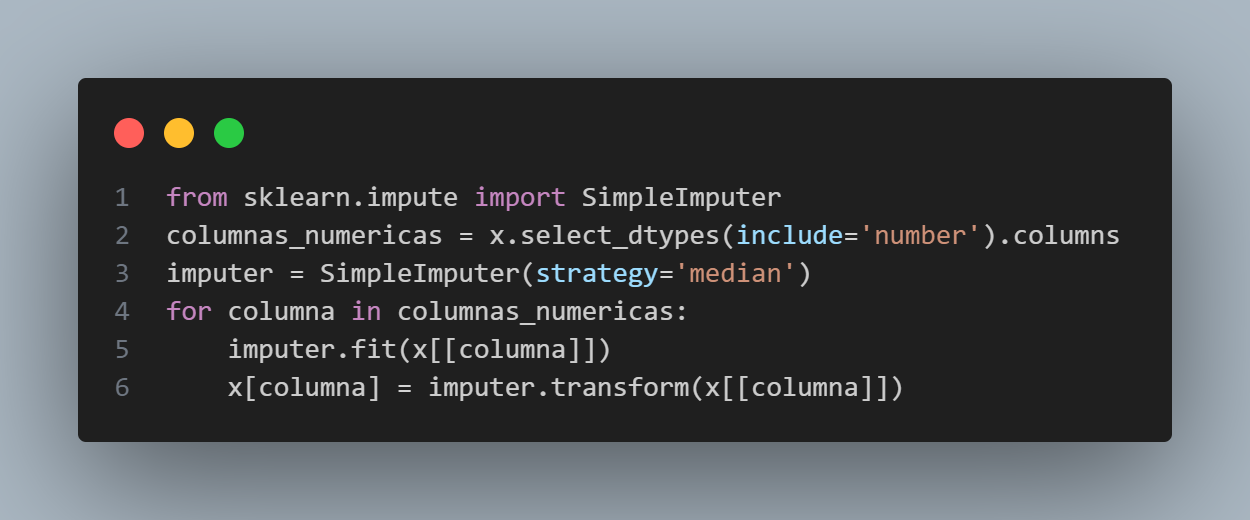


Figura 6. Simple Imputer

**Feature Engineering:**

A raíz de obtener los datos adecuados, era necesario realizar un ajuste a las escalas de estos, debido a que se trataban de diferentes variables. En Python existen varias clases para realizar el escalamiento: *Standard, Min-max scaled, Quantile scaled, Max-abs scaled, Power Scaled* y *Robust scaled*. Todas las clases provienen del módulo sci-kit learn en su sección de *preprocessing*. En la Figura 7 se puede evidenciar de manera gráfica el cambio de escala de los datos utilizados, incluyendo su escala original.

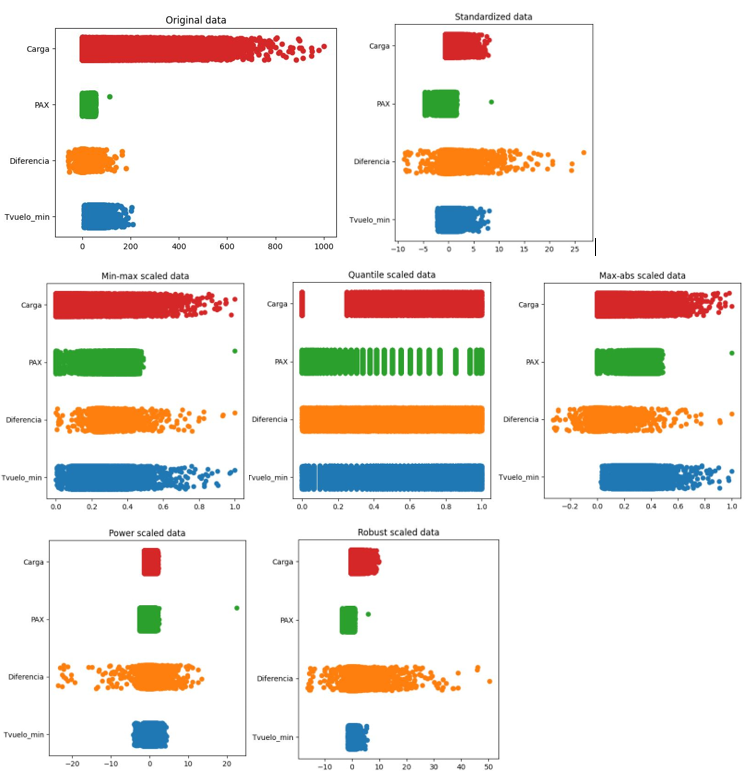


Figura 7. Features aplicados

**Creación del modelo:**

La implementación del algoritmo de Random Forest para nuestros datos se realizó a través de la clase Random Forest Regressor, sin embargo, fue necesario dividir los datos obtenidos en dos: datos de entrenamiento y datos de prueba. Los datos de entrenamiento usualmente representan un 80% del total de los datos disponibles y son estos los que ayudarán a entrenar el modelo de predicción; por otro lado, los datos de prueba son el porcentual restando, en este caso el 20% y con estos se realizarán las pruebas con las métricas de MAE, MSE y RMSE y evidenciar la calidad del modelo. Esta división se realiza con el fin de evitar el *overfitting.*

El *overfitting* es un problema que se puede encontrar al entrenar modelos de predicción supervisados, y radica cuando el modelo memoriza los datos de entrada y su relación con el resultado esperado, mas no aprende a relacionarlos, lo que ocasiona un caos cuando se enfrenta a nuevos datos que no estaban en el *dataset* de entrenamiento (Ying, X., 2019). Un ejemplo de este suceso se evidencia cuando un estudiante de primaria se encuentra en el proceso de aprendizaje de los productos básicos, sin embargo, por diversos factores, no aprende el proceso detrá del producto, el cual es la abreviación de la adición, sino que memoriza los productos y sus multiplos, por lo que cuando se enfrenta a nuevos multiplos o en un diferente orden no será capaz de responder correctamente.

Para evitar este inconveniente se hace uso de la clase *train\_test\_split*, lo que permitirá la división necesaria de los datos. En la Figura 8 se puede evidenciar el uso de esta clase para separar los datos y también el uso de la clase de Random Forest Regressor con una cantidad de árboles de desición indicada por la variable de *n\_estimators*; se obtiene el modelo en el espacio de memoria *regressor* y se hace uso de él con las variables independientes de prueba para generar unos valores dependientes con el modelo desarrollado.

Texto

Descripción generada automáticamente

Figura 8. Random Forest en Python

**Evaluación modelo:**

Finalmente, tras la obtención de las variables dependientes utilizando el modelo generado, se llevó a cabo una evaluación meticulosa de la predicción mediante la implementación de métricas específicas, como el Error Absoluto Medio (MAE), el Error Cuadrático Medio (MSE), y la Raíz del Error Cuadrático Medio (RMSE). Para llevar a cabo esta evaluación, se importó la clase *metrics* del módulo sci-kit learn. La Figura 9 ilustra el proceso de aplicación de estas funciones de evaluación, las cuales toman como entrada los valores de prueba y los valores pronosticados por el modelo.

Texto

Descripción generada automáticamente

Figura 9. Métricas en Python

**Resultados**

Tras llevar a cabo el entrenamiento del algoritmo de aprendizaje supervisado destinado a la predicción del consumo de combustible para vuelos regionales en una aerolínea colombiana, se procedió a una evaluación meticulosa del desempeño del modelo en cada enfoque de ingeniería de características implementado. La evaluación fue diseñada para medir la precisión y eficacia del modelo en la predicción de variables complejas, reflejando así la aplicabilidad de diferentes técnicas de ingeniería de características. Una comparación detallada entre los enfoques se presenta en la Tabla 2, permitiendo un análisis riguroso y profundo de cómo cada método contribuyó al rendimiento global del modelo. Este análisis comparativo es vital para identificar las técnicas más prometedoras y orientar futuras investigaciones y desarrollos en la predicción eficiente del consumo de combustible en el contexto de la aviación regional en Colombia.

Tabla 2. Métricas finales

|  |  |  |  |
| --- | --- | --- | --- |
| **Feature** | **MAE** | **MSE** | **RMSE** |
| Standard Scaled | 23.9867 | 1319.0599 | 36.3188 |
| Min-max Scaled | 23.5430 | 1286.4400 | 35.8669 |
| Quantile Scaled | 23.5402 | 1286.1340 | 35.8627 |
| Max-abs Scaled | 23.5425 | 1286.3764 | 35.8660 |
| Power Scaled | 23.5420 | 1288.4789 | 35.8953 |
| Robust Scaled | 23.5037 | 1285.0845 | 35.8480 |

En la tabla anterior es posible evidenciar seis técnicas de escalado utilizadas: Standard scaled, Min-max scaled, Quantile scaled, Max-abs scaled, Power scaled y Robust scaled. Las anteriores se utilizaron como proceso de la ingeniería de características implementado en los datos de entrenamiento para el algoritmo de *machine learning* generado.

El rendimiento de cada técnica fue evaluado meticulosamente a través de tres métricas comúnmente empleadas en la valoración de algoritmos de pronóstico cuantitativos: el Error Absoluto Medio (MAE), el Error Cuadrático Medio (MSE) y la Raíz del Error Cuadrático Medio (RMSE). Estas métricas proporcionan una medida numérica cuantificable de la precisión de los valores generados.

De las técnicas evaluadas, la denominada "Robust Scaled" se destacó por su excepcional rendimiento, superando a las demás en términos de valores numéricos en todas las métricas analizadas. Con un MAE de 23.5037, un MSE de 1285.0845 y un RMSE de 35.8480, este enfoque demostró una notable capacidad para generar predicciones que se aproximan estrechamente al consumo real de combustible en vuelos regionales de una aerolínea comercial colombiana.

Los valores derivados de estas métricas revelan que el modelo con el enfoque "Robust Scaled" tiene un error promedio de aproximadamente 23 kg de combustible, y un error de hasta 36 kg de combustible en el peor de los escenarios. Estos resultados son particularmente alentadores, ya que señalan un avance significativo en la aplicación de la inteligencia artificial para la predicción precisa del consumo de combustible en la industria aérea colombiana. Adicionalmente, la implementación de este modelo podría traducirse en un manejo más óptimo y eficiente del combustible por parte de las compañías aéreas, resultando en una disminución notable en los gastos totales que habitualmente se enfrentan. En última instancia, estos hallazgos no solo representan un avance tecnológico en la industria aérea, sino que también contribuyen al desarrollo sostenible y a la eficiencia económica en la República de Colombia.

**Discusión**

En relación con los resultados obtenidos a través de las métricas de los diferentes *features*, se puede denotar hay un mejor rendimiento en el *Robust Scaled*. El escalamiento de los datos que se presentan en este se produce al eliminar la media y escala los datos de acuerdo con rangos de los cuantiles, de manera predeterminada, el rango es entre el primer cuartil y el tercer cuartil, representando el veinticincoavo cuantil y el setentaicincoavo cuantil (Pedregosa P., et al., 2011). El escalamiento permite evadir el sesgo que se presenta con valores extremos de los datos, esto ocurre al eliminar la media o promedio, y permite que el valor de los datos cerca a la distribución tengan un mayor impacto.

La aplicación del enfoque Robust Scaled en los datos de entrenamiento para el modelo de predicción del consumo de combustible resultó en una mejora apreciable, con un aumento cercano al 0.4994% en las métricas de evaluación en comparación con otros métodos empleados. Es relevante destacar que este escalamiento o enfoque aplicado a los datos contribuyó a atenuar el sesgo y, por lo tanto, a neutralizar los efectos adversos que los valores atípicos pudieran tener en modelos de predicción fundamentados en datos de entrenamiento.

No obstante, los resultados de las métricas no resultaron óptimos, evidenciando un desajuste aproximado del 25% en relación con los datos reales. Este desajuste podría ser indicativo de la influencia de diversos factores. Para optimizar el modelo de predicción, se requiere un análisis detallado y riguroso, que abarque desde la inclusión de datos faltantes hasta la evaluación de la cantidad de árboles de decisión utilizados en el modelo, y también la consideración de la cantidad de datos de entrenamiento disponibles. Un proceso exhaustivo de identificación y resolución de problemas (troubleshooting) es esencial para detectar posibles desafíos y áreas susceptibles de mejora.

Una estrategia prometedora podría ser procurar una mayor cantidad y calidad de datos para potenciar el rendimiento del modelo. Esto podría requerir la recopilación de datos adicionales en diferentes períodos de tiempo o el perfeccionamiento de los métodos de recolección existentes. Incrementar la cantidad de datos de entrenamiento podría facilitar que el modelo aprenda patrones más precisos y, en consecuencia, realice predicciones más exactas.

Adicionalmente, la evaluación y ajuste de los hiperparámetros del modelo, como la profundidad de los árboles de decisión, podría conducir a la identificación de una configuración óptima que minimice el error en las predicciones. Un calibrado adecuado de estos parámetros podría favorecer la mejora del rendimiento del modelo y la reducción de los valores de las métricas de evaluación. En última instancia, se podría contemplar la utilización de otro modelo de predicción, la incorporación de diferentes características o incluso un replanteamiento del enfoque del estudio, reconociendo que no todas las variables pueden ser pronosticadas eficazmente mediante algoritmos de Machine Learning.

**Conclusiones**

La aplicación de algoritmos de *Machine Learning* para predecir el consumo de combustible en vuelos regionales de una aerolínea colombiana ha tenido un impacto significativo en múltiples aspectos.

En términos de planificación estratégica empresarial, esta predicción precisa del consumo de combustible puede brindar a la aerolínea la capacidad de tomar decisiones informadas en cuanto a la compra de combustible. Al conocer de antemano el consumo estimado en cada vuelo, la aerolínea puede optimizar la gestión de su inventario de combustible, evitando compras innecesarias o insuficientes. Esto conduce a una mejor administración de recursos y una mejora en las utilidades de la compañía.

Además de los beneficios económicos, la predicción del consumo de combustible ha demostrado ser una estrategia efectiva para reducir los costos operativos. Al evitar vuelos vacíos o subutilizados, la aerolínea puede maximizar la eficiencia de sus operaciones y reducir los gastos asociados con vuelos innecesarios. Esto se traduce en ahorros significativos a largo plazo y una mayor rentabilidad para la empresa.

En términos de sostenibilidad ambiental, esta aplicación de *Machine Learning* tiene un impacto positivo al reducir la emisión de gases de efecto invernadero. Al evitar vuelos vacíos, se disminuye la quema innecesaria de combustible, lo que se traduce en una menor huella de carbono de la aerolínea. Esto contribuye al cuidado del medio ambiente y alinearse con los esfuerzos globales para combatir el cambio climático.

Además, esta implementación de algoritmos de *Machine Learning* y el uso de tecnologías avanzadas fortalece el ecosistema de la industria-aeronáutica-gobierno-academia en Colombia. Al adoptar nuevas herramientas y enfoques en la industria aeronáutica, se promueve la innovación tecnológica y se impulsa el desarrollo de capacidades locales en inteligencia artificial y análisis de datos. Esto a su vez tiene un efecto positivo en la economía del país, generando oportunidades de crecimiento y empleo en el sector tecnológico.

En resumen, la aplicación de un modelo predictivo de consumo de combustible basado en *Machine Learning* ha generado beneficios significativos para una aerolínea colombiana. Desde una mejora en la planificación estratégica y la reducción de costos operativos, hasta el cuidado del medio ambiente y el impulso de la innovación tecnológica, esta aplicación representa un avance importante en la industria aeronáutica y su intersección con la inteligencia artificial.

**Referencias**

Baevski, A., Hsu, W., Xu, Q., Babu, A., Gu, J. & Auli, M. (2022). data2vec: A General Framework for Self-supervised Learning in Speech, Vision and Language. Proceedings of the 39th International Conference on Machine Learning, en Proceedings of Machine Learning Research 162:1298-1312 Recuperado de <https://proceedings.mlr.press/v162/baevski22a.html>

Espinosa-Zúñiga, Javier Jesús. (2020). Aplicación de algoritmos Random Forest y XGBoost en una base de solicitudes de tarjetas de crédito. Ingeniería, investigación y tecnología, 21(3), 00002. Epub 02 de diciembre de 2020.<https://doi.org/10.22201/fi.25940732e.2020.21.3.022>

Fabian Pedregosa, Gaël Varoquaux, Alexandre Gramfort, Vincent Michel, Bertrand Thirion, Olivier Grisel, Mathieu Blondel, Peter Prettenhofer, Ron Weiss, Vincent Dubourg, Jake Vanderplas, Alexandre Passos, David Cournapeau, Matthieu Brucher, Matthieu Perrot, Édouard Duchesnay; 12(85):2825−2830, 2011.

Gallet A., Rigby S., Tallman T. N., Kong X., Hajirasouliha I., Liew A., Liu D., Chen L., Hauptmann A. and Smyl D. (2022). Structural engineering from an inverse problems perspective Proc. R. Soc. A.4782021052620210526, <http://doi.org/10.1098/rspa.2021.0526>

Girón Girón, L. Á (2023). Estimación de demanda de transporte aéreo de pasajeros desde un enfoque de redes para aeropuertos tipo hub: caso Aeropuerto El Dorado Bogotá (tesis de maestría). Universidad Nacional de Colombia, Bogotá D.C., Colombia.

Hodson, T. O.: Root-mean-square error (RMSE) or mean absolute error (MAE): when to use them or not, Geosci. Model Dev., 15, 5481–5487, <https://doi.org/10.5194/gmd-15-5481-2022>, 2022.

Janiesch, C., Zschech, P. & Heinrich, K (2021). Machine learning and deep learning. Electron Markets 31, 685–695 (2021). <https://doi.org/10.1007/s12525-021-00475-2>

M. Panda, A. A. A. Mousa and A. E. Hassanien (2021), "Developing an Efficient Feature Engineering and Machine Learning Model for Detecting IoT-Botnet Cyber Attacks," en IEEE Access, vol. 9, pp. 91038-91052, 2021, doi: 10.1109/ACCESS.2021.3092054

Márquez Díaz, Jairo. (2020). Inteligencia artificial y Big Data como soluciones frente a la COVID-19. Revista de Bioética y Derecho, (50), 315-331. Epub 23 de noviembre de 2020. Recuperado en 13 de julio de 2023, de <http://scielo.isciii.es/scielo.php?script=sci_arttext&pid=S1886-58872020000300019&lng=es&tlng=es>

Morrison, J., Bonnefoy, P., Hansman, R. J., & Sgouridis, S. (2011). Recuperado de <http://hdl.handle.net/1721.1/62860>

Orellano Lasprilla, J. L. (2023). Análisis Fluido Dinámico Computacional de un mezclador estático para combustible de aviación JET A-1 y Biodiesel (tesis de maestría). Universidad ECCI, Bogotá D.C., Colombia.

Pizlo, Z. (2001). Perception viewed as an inverse problem. Vision Research, 41(24), 3145–3161. doi:10.1016/s0042-6989(01)00173-0

Schröer, C., Kruse, F., & Gómez, J. M. (2021). A systematic literature review on applying CRISP-DM process model. Procedia Computer Science, 181, 526–534. https://doi.org/10.1016/J.PROCS.2021.01.199.

Viñas, L. C., & Gómez, M. (2023). El impacto de la COVID-19 en la demanda turística internacional. ININEE CIENCIA, 1(1), 23-38.

Ying, X. (2019). An Overview of Overfitting and its Solutions. Journal of Physics: Conference Series, 1168, 022022. doi:10.1088/1742-6596/1168/2/022022

**Fichero aparte**

* Juan Andrés Bermúdez Gómez

Est. Ingeniería Aeronáutica – Escuela de Aviación del Ejército, 2023. Desarrollador *backend junior* de Alura Latam – Rio de Janeiro, Brasil. Integrante semillero de investigación IA y semillero SIE Aviónica de ESAVE. Cel: 3142796747. Correo: [juanbermudez\_10@ieee.org](mailto:juanbermudez_10@ieee.org).

* Cristian Lozano Tafur

Ingeniero aeronáutico de la Fundación Universitaria Los Libertadores en el año 2014, magister en materiales y procesos de la Universidad Nacional de Colombia - Bogotá 2017 y estudiante de doctorado en informática en UNADE - México, docente e investigador de la Fundación Universitaria Los Libertadores y PHC de la Escuela de Aviación del Ejército. [ctistianlozanotafur@cedoc.edu.co](mailto:ctistianlozanotafur@cedoc.edu.co), Cel.3124516847.

* Sara Valentina González Medina

Est. Ingeniería Aeronáutica – Escuela de Aviación del Ejército (ESAVE), 2023. Est. Ingeniería Aeronáutica de sexto semestre de la escuela de Aviación del Ejército (actualmente). Integrante semillero IA. Correo: [valentinagonzalezmedina@cedoc.edu.co](mailto:valentinagonzalezmedina@cedoc.edu.co). Cel. 3152116520.

* Samuel Fuentes Rodríguez

Est. Ingeniería Aeronáutica – Escuela de Aviación del Ejército (ESAVE), 2023. Est. Ingeniería Aeronáutica de sexto semestre de la escuela de Aviación del Ejército (actualmente). Integrante semillero IA. Correo: [samuelfuentesrodriguez@cedoc.edu.co](mailto:samuelfuentesrodriguez@cedoc.edu.co). Cel. 3022109513.

* Pedro Fernando Melo Daza

Ingeniero electrónico de la Universidad Nacional de Colombia – Manizales en el año 2009, especialista en mantenimiento industrial y magister en ingeniería de la Universidad EAFIT – Medellín 2011 y estudiante de doctorado de informática en UNADE – México, Docente tiempo completo de la Fundación Universitaria Los Libertadores y PHC de la Escuela de Aviación del Ejército, investigador asociado ante Min ciencias año 2021, inventor de “System for detecting defects in the roundness of railway vehicle wheels” con registro de patente internacional No. WO2014162270A2, [pfmelod@libertadores.edu.co](mailto:pfmelod@libertadores.edu.co), Cel.3147335921.