Improving fuel control in a Transport Company Using IOT and Machine Learning

Jazmin Madeleine Olivares Quispe; José Roberto Arróspide Ponce; Nicolas Francisco Salazar Medina

Universidad de Lima, Perú, *20194378@aloe.ulima.edu.pe, nsalazar@ulima.edu.pe*
Universidad de Lima, Perú, *20193807@aloe.ulima.edu.pe,*

Abstract–

Fuel management is critical for transportation companies as it is the largest component of operating costs. The company's performance reached an average of only 7.32 km/gal, with the theoretical standard being 9.46 km/gal, having a gap of 2.10 km/gal, this is due to traditional manual measurement methods . since they do not provide accurate or updated data, which prevents making correct decisions regarding the optimal use of fuel. This study proposes the integration of IoT technologies (sensors) and Machine Learning to optimize fuel control and management in a Peruvian company dedicated to heavy cargo transportation, allowing the detection of anomalous patterns in fuel consumption, increasing by 9.1%. gallon yield and reducing fuel consumption by 9.8% in the improved model and a cost savings of 13.8%.

.

Keywords-- IoT, Machine Learning, fuel management

Improving fuel control in a Transport Company Using IOT and Machine Learning

Jazmin Madeleine Olivares Quispe; José Roberto Arróspide Ponce; Nicolas Francisco Salazar Medina

Universidad de Lima, Perú, *20194378@aloe.ulima.edu.pe, nsalazar@ulima.edu.pe*
Universidad de Lima, Perú, *20193807@aloe.ulima.edu.pe,*

***Resumen*– La gestión del combustible es crítica para las empresas de transporte, ya que es el mayor componente de los costos operativos. El rendimiento de la empresa alcanzaba un promedio de solo 7,32 km/gal, siendo el teórico estándar de 9,46 km/gal, teniendo una brecha de 2,10 km/gal, esto se debe a los métodos tradicionales de medición manual ya que no brindan datos precisos ni actualizados, lo que impide tomar decisiones acertadas en el uso óptimo del combustible. Este estudio propone la integración de tecnologías IoT (sensores) y Machine Learning para optimizar el control y gestión del combustible en una empresa peruana dedicada al transporte de carga pesada, permitiendo detectar patrones anómalos en el consumo de combustible, incrementando en 9,1% de rendimiento de galones y disminuyendo el consumo del combustible en 9,8% en el modelo mejorado y un ahorro en costos del 13,8%.**

Palabras clave-- IoT, Machine Learning, gestión de combustible

I. INTRODUCCIÓN

El sector transporte en Perú ha enfrentado una crisis debido a la pandemia del COVID-19, lo que ocasionó alzas en los precios de hidrocarburos y afectó la logística nacional. Según el INEI (2022), durante el último año, el diésel incrementó en 21.5%, el gasohol en 13.9% y el gas natural en 3.3% [1]. A pesar de estas dificultades, el sector transporte logró un aumento del 10.2% en su valor agregado bruto en 2022, posicionándose como el segundo sector con mayor crecimiento [2].

Uno de los grandes problemas en el sector de transporte de carga es la falta de exactitud en el control de combustible. Por ello, se propone el uso de nuevas tecnologías de recolección y monitoreo que sean más precisas y convenientes, además de ofrecer diferentes medios de extracción [3].

Estudios recientes han señalado que el uso de tecnologías como IoT, específicamente el empleo de escáneres, mejora significativamente la precisión en el monitoreo de combustible. Además, los modelos de machine learning han demostrado su eficacia para el análisis predictivo del consumo y la detección de patrones anómalos basados en datos recopilados por sistemas IoT.

En este contexto, la empresa seleccionada, dedicada al transporte y abastecimiento de combustible para grifos, enfrenta un problema crítico relacionado con la inexactitud en la medición y el control del consumo de combustible. Actualmente, sus métodos tradicionales de medición manual no proporcionan datos precisos ni actualizados, lo que afecta su capacidad para tomar decisiones informadas y optimizar el uso del combustible.

Por ello, este estudio propone una mejora en el control del consumo de combustible mediante la implementación de tecnologías de IoT y algoritmos de machine learning. La solución requiere instalar un escáner J1939 en cada camión para extraer y exportar datos a la nube, y luego utilizar esos datos para desarrollar un modelo predictivo que permita detectar consumos anómalos. Se plantea como hipótesis que el uso combinado de IoT y ML mejora significativamente la gestión del combustible. Este problema es crucial, ya que actualmente el rendimiento promedio de la empresa es de solo 7.32 km/gal frente al estándar del sector de 10.8 km/gal, mostrando una brecha considerable de 3.48 km/gal. Además, el combustible representa el 82.8% de los costos operativos de la empresa, por lo que una gestión más eficiente no solo generaría ahorros económicos, sino también contribuiría a la sostenibilidad ambiental al reducir hasta en un 25% las emisiones de gases de efecto invernadero [4][5].

II REVISIÓN DE LA LITERATURA

Este estudio analizó sistemas de control de combustible para transporte de carga pesada mediante una Revisión Sistemática de Literatura (RSL) en Scopus y Web of Science, siguiendo la metodología PRISMA. De 1780 documentos

iniciales (2018-2022), se seleccionaron 40 artículos relevantes, destacando contribuciones de China, India, Reino Unido y Suecia en tecnologías como IOT (45%), Machine Learning (30%) y modelos de regresión logística (10%), para optimizar el consumo de combustible. Se evaluaron dos sistemas de telemetría, optando por un escáner OBD2 con conectividad GSM, que recopila datos procesados en la nube mediante un modelo SVM, asegurando un monitoreo eficiente y preciso.

II DEFINICIÓN DEL PROBLEMA

La empresa enfrenta un problema significativo de gestión inadecuada del consumo de combustible debido a métodos manuales imprecisos, como el uso de varillas, y la falta de automatización en los procesos. Esto genera errores en las mediciones, dependencia de una sola persona para realizar las tareas y una ausencia de control en tiempo real, lo que impide detectar consumos anómalos a tiempo.

Decidiéndose utilizar machine learning para detección de regularidades en el consumo, para registros automáticos y predicciones del consumo, mientras que el IOT para obtener datos precisos con respecto a la medicación.

III METODOLOGÍA

El tamaño de muestra fue determinado utilizando la fórmula de población finita de Spiegel y Stephens (2009), considerando un nivel de confianza del 90% y un error admisible de 0.1, obteniendo una muestra representativa de 15 camiones[6]. La variable principal considerada fue el consumo de combustible.

Se excluyeron los grifos ubicados en provincias fuera del área de estudio y se evitó la recolección de datos en días feriados, ya que en estos días ocurren descargas atípicas. Esto aseguró que los datos recolectados fueran representativos de la operación regular.

Para estimar el número adecuado de réplicas para la simulación, se empleó el método Half-Width, obteniéndose 715 réplicas con un intervalo de confianza del 90% y un margen de error de 0.1.

Fig. 1 Esquema de flujo de información

Fig. 2 Scanner utilizado para la instalación

Una vez con los datos obtenidos del scanner, que se detallaran en los componentes, se procederá la creación de modelo con el algoritmo de Support Vector Machine (SVM) que predice un valor específico o un conjunto de clases, ya sea en forma de clasificación o regresión [7]. Aunque este método obtuvo un R-cuadrado 0,96, en este estudio deseamos comparar múltiples algoritmos para seleccionar el más preciso como se hizo en un estudio de Canadá, con el cual obtuvieron una precisión de con un modelo de machine learning de GBM con una diferencia media de 1,70% [8].

Estas predicciones de consumos son extremadamente útiles para la gestión eficiente de los recursos, sin embargo, también requerimos detectar los comportamientos anómalos de niveles de combustibles. Por ende, se optó por el uso de un Box and Whisker Plot para la detección de consumos anómalos, los cuales pueden ser explicados por modelo de Machine Learning (ML) previamente entrenado

Se emplearon datos recogidos por sensores a intervalos de 2 minutos durante todo el trayecto del camión cisterna, desde la salida del estacionamiento inicial, donde se mide el combustible por primera vez, hasta su llegada a la zona 4, donde se realiza la última medición del combustible y se guarda el camión cisterna Estos sensores registran información sobre kilómetros, velocidad, nivel de combustible, consumo y RPM del motor; estos datos

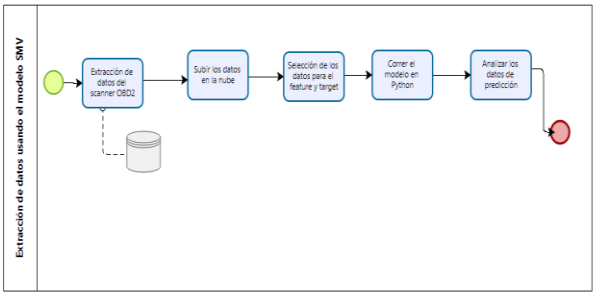


Fig. 3 Proceso de Flujo de extracción de datos después de la implementación del modelo SMV

Para validar el modelo se utilizó el método Mínimo Cuadrado Ordinario (OLS), se validó que estas variables tienen un impacto significativo en el consumo de combustible, ya que los valores p-value para cada una fueron menores a 0.05 y además el índice de correlación para ver cómo es la relación.

Obteniendo lo siguiente: los kilómetros recorridos, la

velocidad y las RPM tienen una correlación positiva, es decir; que el consumo de combustible aumenta a medida que estos factores se incrementan. Mientras que el consumo de combustible y nivel de combustible tienen una relación indirectamente proporcional es decir, el consumo de combustible es mayor su nivel de combustible es menor una de las razones se debe a que cuando el tanque está más lleno se tiene menos espacio para que se evapore según la revista la Dirección General de Tráfico, DGT ” en su edición conducción eficiente del año 2023.

Adicionalmente, se ha decidido hallar el Mean Squared Error (MSE) para evaluar la precisión de la predicción, similar a uno utilizado en un estudio realizado en Singapur en donde se menciona lo siguiente: un MSE bajo indicaría que el modelo está haciendo predicciones cercanas a la realidad, lo que es esencial para la interpretación y utilidad del modelo en la aplicación práctica; además durante el proceso se puede hacer ajuste de variables para aumentar la precisión. [9].

Los resultados de Mean Squared Error (MSE) es de 0.0002 y un R^2 de 0.9433, lo cual significa que el modelo está haciendo predicciones cercanas a la realidad y explica aproximadamente el 94.3% de la variabilidad en los datos del consumo del combustible.

Para la modelación y simulación se utilizó el software Arena 16.20, en el modelo propuesto, no se modifica el flujo; sin embargo, se cambiarán los datos. En el primer modelo, los datos se basan en información manual, mientras que en el modelo propuesto no se considera la actividad manual de medición de combustible. En su lugar, se utilizará el nuevo consumo promedio generado por el modelo SVM con el modelo óptimo previamente entrenado.

La variable de principal interés de este modelo de simulación son los consumos de combustible que se obtienen en el modelo inicial y mejora.

Por ende, con el fin de realizar la comparación entre ambos escenarios es necesario el modelo de machine learning para obtener el consumo optimizado.

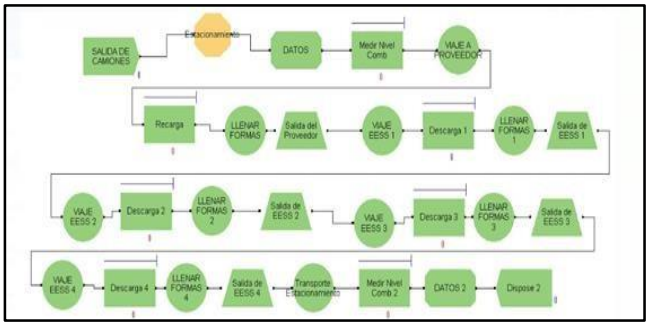


Fig. 4: Modelo de recorrido de la empresa en el software Arena

Además, en el modelo propuesto se espera evitar las anomalías o, en su defecto, detectar cualquier irregularidad de manera temprana, lo que permitirá reducir los costos operativos y, además, mejorar la precisión en el cálculo del consumo necesario para lograr un rendimiento cercano al estándar del sector, que es de 10,8 km. En el caso del modelo desarrollado, si se detecta alguna anomalía, el sistema informa el número de registro donde ocurrió y proporciona todos los factores utilizados en el modelo para su análisis. A continuación, se presentarán las gráficas que ilustran tanto la presencia de datos atípicos como la ausencia de irregularidades en los datos

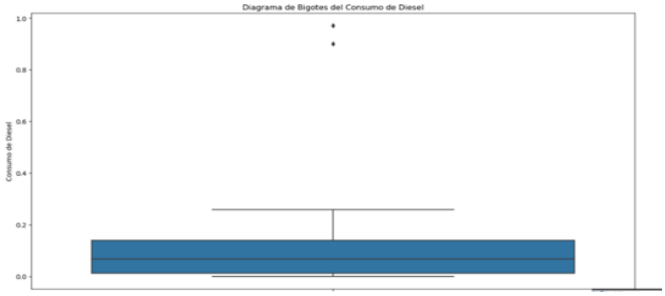


Fig 5: Gráfico de cajas del método SMV presentando anómalos

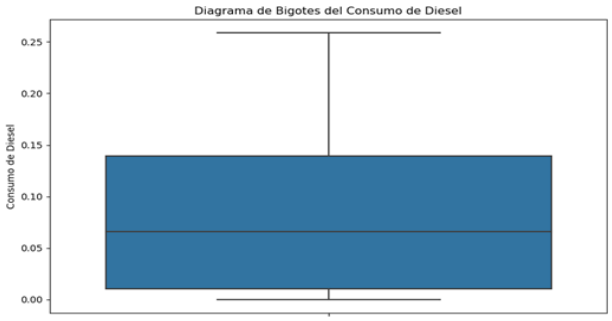


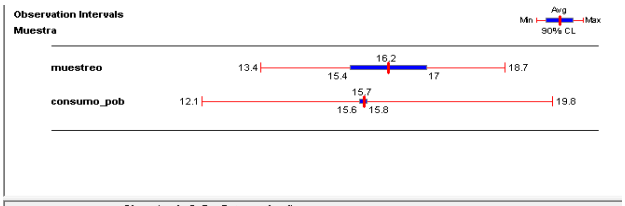
Fig 6: Gráfico de cajas del método SMV sin presentar anómalos

Dentro del marco teórico, se incluyó un análisis detallado de las técnicas estadísticas aplicadas en estudios previos, especialmente en cuanto a la validación de modelos predictivos y el uso de machine learning en la gestión de flotas. Las estadísticas obtenidas en este estudio, como el R^2 y el MSE, son consistentes con los resultados de la literatura, lo que valida la efectividad del enfoque adoptado. Esta base teórica refuerza la importancia de utilizar un enfoque multidimensional para el análisis del consumo de combustible, integrando variables como el comportamiento del conductor, las condiciones de la ruta y el estado del vehículo, todos ellos reflejados en los modelos de machine learning.

IV RESULTADOS

Con el fin de comparar el modelo base y modelo propuesto, se han analizado los resultados de los intervalos de confianza de los indicadores en ambos escenarios.

En los siguientes gráficos se presentan los rangos de cada intervalo para identificar posibles superposiciones. En caso de encontrarse una superposición, se procede a realizar la prueba pareada en el Output Analyzer para determinar si la mejora es significativa



Classical C.I. Intervals Summary
Muestra

IDENTIFIER	AVERAGE	STANDARD DEVIATION	0.900 C.I. HALF-WIDTH	MINIMUM VALUE	MAXIMUM VALUE	NUMBER OF OBS.
muestreo	16.2	1.46	0.845	13.4	18.7	10
consumo_pob	15.7	1.57	0.097	12.1	19.8	715

Fig 7: Grafico de medias de Output Analizer

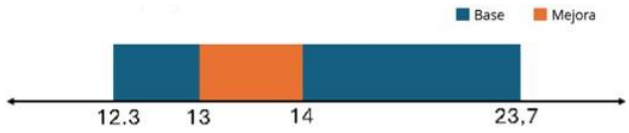


Fig 8:Galones Consumidos en ambos escenarios

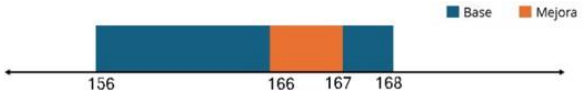


Fig 9: Costos en ambos escenarios

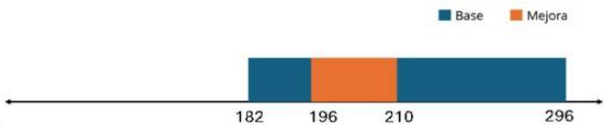


Fig 10: Restante de combustible en ambos escenarios

El precio del combustible utilizado para los cálculos es el diésel B5 S-50 UV, ya que todos los camiones emplean este tipo de diésel. Para los primeros cuatro meses, se tomaron los precios promedio proporcionados por Osinergmin (2023), que fueron 18.35, 17.49, 16.43 y 15.59 soles por galón, respectivamente. [10]

Para los meses posteriores, debido a la falta de información de Osinergmin, se utilizó el precio de 15.79 soles por galón de la fuente [11].Se decidió emplear un precio promedio de S/ 15 por galón para los cálculos.

A partir de la simulación en Arena y el modelo SMV, se logró un aumento del 9.1% en el rendimiento de los galones y una disminución del 9.8% en el consumo de combustible

TABLA I
COMPARACIÓN DEL MODELO ANTES Y DESPUÉS

	ilómetro	Consumo de combustible (Galones)	Costo	Rendimiento (kilómetros/galones)
Modelo Base	110,28	15,05	S/ 225,75	7,32
Modelo Mejorado	110,28	13,57	S/ 203,55	8,12

V DISCUSIÓN

Se comparan los resultados del modelo elaborado con los de los artículos mencionados en la revisión de la literatura.

TABLA 2:
COMPARACIÓN DEL MODELO Y ARTÍCULOS REVISADOS

	BENCHMARK		
	Resultado del modelo	VS_MAF	Gradient Boost
R-CUADRADO	0,964	0,97	0,77

El modelo VS MAF utiliza dos variables: velocidad y flujo másico de aire. Su resultado elevado de R^2 es ligeramente superior, ya que ellos determinaron, mediante algoritmos de Random Forest y árboles de decisión, que estas variables tienen una mayor influencia en el consumo [8].

Por otro lado, el modelo de Gradient Boost [12], obtuvo un resultado inferior, ya que la frecuencia con la que se extraían los datos era de 10 minutos, mientras que la nuestra fue de 2 minutos. Además, no se realizaron muestras estandarizadas, ya que cada viaje analizado podía ser diferente al anterior.

En este estudio, los resultados obtenidos con el modelo propuesto se comparan con varios trabajos previos en la literatura, con el fin de situar nuestro modelo en el contexto de investigaciones similares. Nuestro modelo alcanzó un R^2 de 0,964, lo cual es comparable con el rendimiento de otros modelos avanzados. Por ejemplo, el estudio de Heni [8] empleó un modelo de Gradient Boosting con un R^2 de 0,94, lo que es ligeramente inferior al rendimiento de nuestro modelo. La diferencia en los resultados podría explicarse por las variaciones en los datos de entrada: mientras que en nuestro caso los datos se recopilaron a una frecuencia de 2 minutos, en el estudio de Heni [8] los datos fueron tomados cada 10 minutos. Esta diferencia en la frecuencia de muestreo probablemente mejora la precisión de las predicciones de nuestro modelo, ya que se captura un nivel de detalle más fino de las variaciones en el consumo de combustible.

Por otro lado, el modelo VS MAF [8], que utiliza variables como la velocidad y el flujo másico de aire, reportó un R^2 de 0,97, que es superior al de nuestro modelo en cuanto a precisión. Sin embargo, el modelo VS MAF empleó algoritmos como Random Forest y árboles de decisión, los cuales tienen una mayor capacidad para identificar

interacciones no lineales entre variables, lo que puede explicar su rendimiento ligeramente superior. A pesar de esto, nuestro modelo con SVM tiene la ventaja de ser más flexible en la generalización de patrones y en la capacidad de manejar datos con características complejas, como las que se presentan en operaciones de transporte en tiempo real.

En términos de eficiencia de combustible, los resultados obtenidos en nuestro estudio muestran una mejora del ~9,8% en el consumo de combustible, lo cual es consistente con los hallazgos reportados en estudios recientes de eco-conducción asistida por IA. Por ejemplo, Chen [13]reportaron una reducción del consumo de combustible de aproximadamente el 15%, mientras que Liu [14] lograron mejoras sustanciales utilizando deep reinforcement learning en escenarios urbanos. Sin embargo, en nuestro caso, el ahorro obtenido es menor, lo cual es comprensible dado que nuestro enfoque estuvo centrado más en el monitoreo de datos y la detección de anomalías, en lugar de una optimización activa de las rutas o la conducción en sí. Este tipo de enfoque más conservador en la optimización ha sido documentado en otros estudios, como el de Fafoutellis[15], donde los ahorros de combustible en la implementación de tecnologías de eco-conducción asistida por IA fueron del orden del 10-15%.

La principal contribución de este estudio radica en la implementación de un sistema de telemetría IoT combinado con modelos de machine learning para la predicción y monitoreo en tiempo real del consumo de combustible, lo que mejora considerablemente la precisión en comparación con los métodos tradicionales de control manual. Aunque nuestro modelo no alcanzó los niveles más altos de ahorro de combustible reportados en otros estudios, ofrece un valor agregado significativo en la identificación temprana de anomalías, como fugas o comportamientos de conducción ineficientes, que podrían pasar desapercibidos utilizando métodos convencionales.

En resumen, aunque los resultados de este estudio muestran una mejora significativa en la eficiencia de combustible, estos se encuentran dentro del rango de ahorros típicos reportados en la literatura, lo que valida la efectividad del enfoque. La comparación con otros modelos demuestra que nuestra metodología, aunque no supere todos los estudios previos en términos de R^2 o reducción de consumo, ofrece ventajas claras en términos de monitoreo constante y detección de anomalías, lo que representa una aportación valiosa a la gestión de flotas y la reducción de costos operativos.

VI CONCLUSIONES

Este proyecto evidencia que la implementación de tecnologías de IoT y modelos de machine learning logra una mejora significativa en la precisión y eficiencia del control de consumo de combustible para el sector transporte, comparado con los métodos tradicionales de medición manual, los cuales mostraban imprecisiones y ocasionaban sobrecostos.

Se obtuvo un aumento del 9.1% de rendimiento de galones y una disminución de consumo de combustible del 9.8%, asimismo, alineado con los ODS de la compañía. No obstante, enfrenta desafíos ambientales significativos (ODS 15), especialmente en el manejo de combustibles, con un impacto neto de -124 según la Matriz de Leopold. La implementación del sistema de alertas de consumo de combustible por anomalías (fugas, obstrucción etc.) ayuda a disminuir el impacto de la huella de carbono al igual que el IOT que incrementa la digitalización disminuyendo el uso de papel.

VII RECOMENDACIONES

Se sugiere la exploración de modelos predictivos avanzados o híbridos, que permitan mejorar aún más la precisión de los algoritmos actuales en la detección y predicción de patrones de consumo anómalos. Asimismo, se recomienda extender el monitoreo a otras variables operativas del camión (carga y velocidad), para obtener un análisis integral del consumo de combustible en diversas condiciones de operación. También es aconsejable la implementación de sensores de succión ya que se puede tener una comparación más directa en cuanto al consumo. Finalmente, se propone realizar estudios comparativos en empresas del mismo sector para validar y ajustar la solución en distintos contextos y tipos de vehículos

AGRADECIMIENTO

Deseamos expresar nuestro más sincero agradecimiento a la Universidad de Lima, por habernos brindado las herramientas y conocimientos necesarios para desarrollar este artículo. Agradecemos profundamente a los docentes que han sido fundamentales en nuestro crecimiento académico y profesional.

REFERENCIAS

[1] INEI. (2022). Indice de precios al consumidor variación % de precios al consumidor de lima metropolitana: Diciembre 2022 Fuente: Instituto Nacional de Estadística e Informática.

[2] *Sutran (2022) Economía peruana creció un 2.7% durante 2022: ¿Cuál es el impacto de las protestas y sequías?* . COMEXPERU - Sociedad de Comercio Exterior Del Perú. Spiegel,

[3] Gong, J., Shang, J., Li, L., Zhang, C., He, J., & Ma, J. (2021). A comparative study on fuel consumption prediction methods of heavy-duty diesel trucks considering 21 influencing factors. Energies, 14(23). https://doi.org/10.3390/en1423

[4] Ghaffarpasand, Omid, Mark Burke, Louisa K. Osei, Helen Ursell, Sam Chapman, and Francis D. Pope. 2022. "Vehicle Telematics for Safer, Cleaner and More Sustainable Urban Transport: A Review" Sustainability 14, no. 24: 16386. https://doi.org/10.3390/su142416386

[5] Jiménez, Elías. (2018). Impacto del precio del diésel y otros insumos en los costos de operación del autotransporte de carga. Publicación Técnica No.536. Instituto Mexicano del Transporte. México.

[6] M., & Stephens, L. (2009). Estadística. 4ta edición. In *Mc Graw-Hill. México.*

[7] Hamed, M. A., Khafagy, M. H., & Badry, R. M. (2021). Fuel Consumption Prediction Model using Machine Learning. International Journal of Advanced Computer Science and Applications(Vol. 12, Issue 11). www.ijacsa.thesai.org

[8] Hamza Heni , S. Arona Diop , Jacques Renaud y Leandro C. Coelho (2023) Medición del consumo de combustible en las rutas de vehículos: nuevos modelos de estimación mediante aprendizaje supervisado, International Journal of Production Research, 61:1, 114-130, DOI:10.1080/ 00207543.2021.1948133

[9] Anandamurugan, S., Deenadhayalan, R., Venkatesan, B., Sakthivel, S. y Rajesh,S. (2022). Un método eficiente para predecir el consumo promedio de combustible en automóviles es utilizando el apilamiento de conjuntos en Python. En Avances en ciencia y gestión de datos: Actas de ICDSM 2021 (págs. 197-209). Singapur: Springer Nature Singapur. https://doi.org/10.1007/978-981-16-5685-9_19

[10] Osinergmin. (2023). PRECIOS PROMEDIO DE COMBUSTIBLES (Soles/Galón) EESS/GRIFOS Y GASOCENTROS - MARZO 2023.

[11] GlobalPetrolPrices. (2023). Perú Precios del diesel. https://es.globalpetrolprices.com/Peru/diesel_prices/

[12] Bousonville, T., Dirichs, M. and Krüger, T. (2020). Estimating truck fuel consumption with machinelearning using telematics, topology and weather data. IESM, 1-6. 10.1109/IESM45758.2019.8948175"

[13] Gao, J., Chen, H., Dave, K., Chen, J., Li, Y., Li, T., & Liang, B. (2020). Analysis of driving behaviours of truck drivers using motorway tests. 234(13), 3103–3111. https://doi.org/10.1177/0954407020925568

[14] Yang, Y., Gong, N., Xie, K., & Liu, Q. (2022). Predicting Gasoline Vehicle Fuel Consumption in Energy and Environmental Impact Based on Machine Learning and Multidimensional Big Data. Energies15(5). https://doi.org/10.3390/en15051602

[15] Yang, Y., Gong, N., Xie, K., & Liu, Q. (2022). Predicting Gasoline Vehicle Fuel Consumption in Energy and Environmental Impact Based on Machine Learning and Multidimensional Big Data. Energies15(5). https://doi.org/10.3390/en15051602