



UNIVERSIDAD POLITÉCNICA SALESIANA
SEDE CUENCA
CARRERA DE INGENIERÍA AUTOMOTRIZ

**MODELO PREDICTIVO PARA LA DETECCIÓN DE FALLAS INCIPIENTES MEDIANTE
APRENDIZAJE AUTOMÁTICO Y PID's EN EL SISTEMA DE INYECCIÓN DE UN KIA
SPORTAGE 2.0L**

Trabajo de titulación previo a la obtención del
título de Ingeniero Automotriz

AUTORES: JOSTIN SANTIAGO ALVAREZ MONTENEGRO
VÍCTOR SAÚL SIAVICHAY NEIRA
TUTOR: ING. NÉSTOR DIEGO RIVERA CAMPOVERDE, MSc.

Cuenca - Ecuador

2024

CERTIFICADO DE RESPONSABILIDAD Y AUTORÍA DEL TRABAJO DE TITULACIÓN

Nosotros, Jostin Santiago Alvarez Montenegro con documento de identificación N° 0151005022 y Víctor Saúl Siavichay Neira con documento de identificación N° 0150532141; manifestamos que:

Somos los autores y responsables del presente trabajo; y, autorizamos a que sin fines de lucro la Universidad Politécnica Salesiana pueda usar, difundir, reproducir o publicar de manera total o parcial el presente trabajo de titulación.

Cuenca, 24 de junio del 2024

Atentamente,



Jostin Santiago Alvarez Montenegro

0151005022



Víctor Saúl Siavichay Neira

0150532141

CERTIFICADO DE CESIÓN DE DERECHOS DE AUTOR DEL TRABAJO DE TITULACIÓN A LA UNIVERSIDAD POLITÉCNICA SALESIANA

Nosotros, Jostin Santiago Alvarez Montenegro con documento de identificación N° 0151005022 y Víctor Saúl Siavichay Neira con documento de identificación N° 0150532141, expresamos nuestra voluntad y por medio del presente documento cedemos a la Universidad Politécnica Salesiana la titularidad sobre los derechos patrimoniales en virtud de que somos autores del Artículo académico: “Modelo predictivo para la detección de fallas incipientes mediante aprendizaje automático y PID’s en el sistema de inyección de un Kia Sportage 2.0L”, el cual ha sido desarrollado para optar por el título de: Ingeniero Automotriz, en la Universidad Politécnica Salesiana, quedando la Universidad facultada para ejercer plenamente los derechos cedidos anteriormente.

En concordancia con lo manifestado, suscribimos este documento en el momento que hacemos la entrega del trabajo final en formato digital a la Biblioteca de la Universidad Politécnica Salesiana.

Cuenca, 24 de junio del 2024

Atentamente,



Jostin Santiago Alvarez Montenegro

0151005022



Víctor Saúl Siavichay Neira

0150532141

CERTIFICADO DE DIRECCIÓN DEL TRABAJO DE TITULACIÓN

Yo, Néstor Diego Rivera Campoverde con documento de identificación N° 0103898995, docente de la Universidad Politécnica Salesiana, declaro que bajo mi tutoría fue desarrollado el trabajo de titulación: MODELO PREDICTIVO PARA LA DETECCIÓN DE FALLAS INCIPIENTES MEDIANTE APRENDIZAJE AUTOMÁTICO Y PID's EN EL SISTEMA DE INYECCIÓN DE UN KIA SPORTAGE 2.0L, realizado por Jostin Santiago Alvarez Montenegro con documento de identificación N° 0151005022 y por Víctor Saúl Siavichay Neira con documento de identificación N° 0150532141, obteniendo como resultado final el trabajo de titulación bajo la opción Artículo académico que cumple con todos los requisitos determinados por la Universidad Politécnica Salesiana.

Cuenca, 24 de junio del 2024

Atentamente,



Ing. Néstor Diego Rivera Campoverde, MSc.

0103898995

MODELO PREDICTIVO PARA LA DETECCIÓN DE FALLAS INCIPIENTES MEDIANTE APRENDIZAJE AUTOMÁTICO Y PID's EN EL SISTEMA DE INYECCIÓN DE UN KIA SPORTAGE 2.0L

PREDICTIVE MODEL FOR THE DETECTION OF INCIPIENT FAILURES USING MACHINE LEARNING AND PID CONTROLLERS IN THE INJECTION SYSTEM OF A KIA SPORTAGE 2.0L

Jostin Santiago Álvarez Montenegro¹, Víctor Saúl Siavichay Neira²

Resumen

Este artículo describe un estudio que emplea un enfoque predictivo para analizar el comportamiento de motores de combustión interna en condiciones reales de conducción. El modelo se basa en la recolección de datos mediante el puerto OBD II del vehículo, con respecto al funcionamiento del motor en diversas condiciones de operación. Principalmente, se registran 3 escenarios en el sistema de inyección: funcionamiento normal, filtro obstruido y bujías con una apertura de 1.3 mm, cumpliendo con los parámetros de ruta predeterminados en el RDE (Real Driving Conditions). Inicialmente, se utilizó el algoritmo K-Nearest Neighbors como modelo de aprendizaje automático para entrenar y validar su desempeño en diversas condiciones operativas. Los datos de los Parámetros de Identificación (PID's) más relevantes, determinados mediante Random Forest, fueron indicadores fundamentales para caracterizar el efecto de las fallas provocadas en cada escenario generado. Estos datos, junto con las emisiones de CO₂, CO, NO_x y HC como factores resultantes de cada iteración, conforman la base de datos que alimenta al modelo de predicción. Al final, este modelo presenta una estrategia relevante como herramienta de diagnóstico con una efectividad del 98% para predecir las fallas descritas con efectos directos en las emisiones contaminantes.

Palabras Clave: OBD-II, PID's, K-Nearest Neighbors, RDE emisiones.

Abstract

This article describes a study employing a predictive approach to analyze the behavior of internal combustion engines under real driving conditions. The model relies on data collection via the vehicle's OBD II port, focusing on engine operation across various operational scenarios. Primarily, three injection system scenarios are recorded: normal operation, clogged filter, and spark plugs with a 1.3 mm gap, complying with predefined route parameters in Real Driving Emissions (RDE) conditions. Initially, the K-Nearest Neighbors algorithm was used as a machine learning model to train and validate its performance under diverse operational conditions. The most relevant Parameter Identification (PID) data, determined through Random Forest, were crucial indicators to characterize the effects of faults in each generated scenario. These data, along with CO₂, CO, NO_x, and HC emissions as resulting factors from each iteration, constitute the database feeding the prediction model. Ultimately, this model presents a significant strategy as a diagnostic tool with 98% effectiveness in predicting the described faults, which directly impact pollutant emissions.

Keywords: OBD-II, PID's, K-Nearest Neighbors, RDE emissions

1. Introducción

La emisión de contaminantes atmosféricos por parte del parque vehicular constituye un desafío global, debido a sus severas implicaciones para la salud pública y el medio ambiente [1-2]. La presencia prolongada de estos gases en áreas urbanas impulsa la implementación de normativas ambientales más estrictas y la búsqueda de alternativas para reducir la contaminación, donde un problema común es el desconocimiento del estado de los vehículos y el efecto en las emisiones contaminantes, consumo de combustible y pérdidas de rendimiento [3-4].

El OBD (On Board Diagnostics) es un sistema de diagnóstico incorporado en los vehículos con unidades de control electrónico (ECU). Esta herramienta facilita la capacidad de controlar y monitorear los diferentes sistemas del motor de combustión interna (ICE) y demás apartados [5-6]. La segunda versión de los sistemas de diagnóstico, OBD – II, incluye el control de las emisiones contaminantes y el acceso al análisis de las señales PID's de cada sensor en el vehículo. De esta forma, se abre la posibilidad de nuevas funciones de diagnóstico [7].

En búsqueda de un modelo de detección temprana de fallos mediante los parámetros de cada sensor, es necesario la recopilación de las señales PID's que contengan un factor de diferenciación, esto permitirá resaltar el comportamiento de las señales de los sensores en cada iteración [8]. Iniciando con pruebas de rutas con el motor en un correcto estado y luego de generar defectos como un filtro de aire saturado y bujías con una separación exagerada debido al alto impacto que estas generan en el rendimiento de un motor [9-10]. Así mismo, el producto de cada prueba está reflejado en los principales gases de escape [11-17]. Generando de esta forma una variable que agrega mayor campo de análisis sobre lo que sucede en el interior del motor después de alterar variables de desempeño en el mismo.

En consecuencia, este modelo plantea la integración entre los datos de OBD II recopilados y los algoritmos de aprendizaje automático, esto en respuesta a los modelos ya realizados por medio de análisis de vibración que utilizan múltiples equipos relacionados a su aplicación [8-22], lo que permite generar un algoritmo de detección temprana de fallas para el monitoreo y control proactivo de los sistemas de inyección mediante una propia plataforma del vehículo [12-13].

2. Materiales y Métodos

En esta sección se presenta la metodología empleada en la investigación, así como los materiales y equipos utilizados durante el desarrollo del trabajo.

2.1. Metodología

El diseño experimental se centra en establecer un procedimiento detallado para la adquisición, procesamiento y análisis de datos, como se ilustra en el diagrama de flujo de la Figura 1. La premisa implica recolectar datos de las variables PID's y los niveles de emisiones contaminantes (PEMS) bajo tres condiciones operativas distintas [11], cada una con diferentes configuraciones en el sistema de inyección: condición normal (NF), filtro saturado (F1) y bujías con una apertura de 1,3 mm (F2). Estos conjuntos de datos determinan el comportamiento del motor de combustión bajo las condiciones mencionadas, de forma que, las variables más relevantes de este conjunto se convierten en la principal base para entrenar un modelo capaz de analizar y validar los datos mediante técnicas de aprendizaje automático [14-15]. En última instancia, el objetivo desarrollar un modelo predictivo de fallas en el sistema de inyección que debe suplir como una herramienta de retroalimentación tanto para la anticipación de fallas en el mantenimiento vehicular como reducir de los factores de emisión.

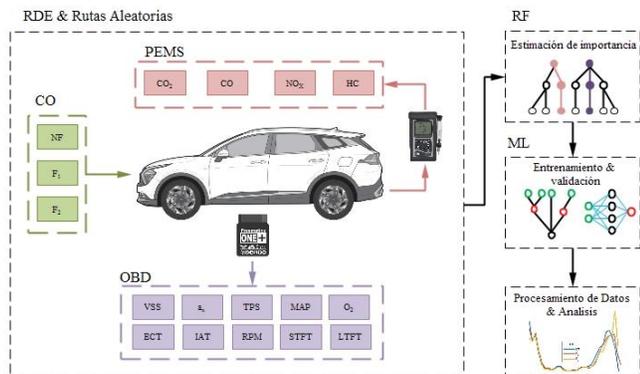


Fig. 1. Procedimiento para la creación del modelo predictivo de fallas.

La recolección de los datos de estudio se da mediante las corridas experimentales que están acorde con los parámetros establecidos por la RDE (Real Driving Emissions) y la normativa europea Euro 6 [18], los segmentos recorridos son 21.63 Km en urbano, 21.24 Km en rural y 19.61 Km en autopista como se observa en la Figura 2, cada uno cumpliendo con los 16 km mínimos para cada segmento. Los datos recolectados comprenden las señales (PID's) de los sensores que comprenden el sistema de inyección del vehículo y la emisiones contaminantes resultantes de las condiciones de funcionamiento del motor y el entorno de la ruta [19].

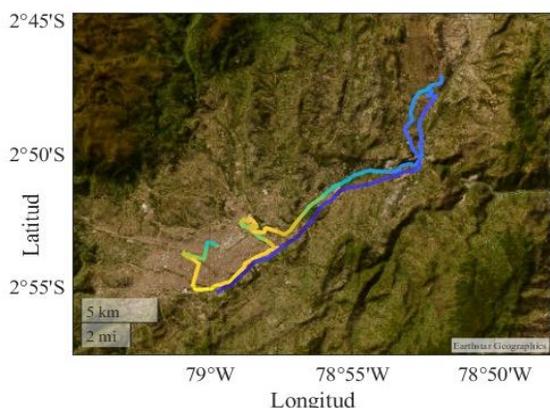


Fig. 2. Ruta RDE para adquisición de datos.

2.1.1. Estimación de emisiones contaminantes

Las medidas de emisiones registradas por el analizador portátil se presentan en términos de concentración volumétrica, desde donde se

determinan los caudales masicos para cada contaminante [11-17]. El primero es el flujo masico \dot{m}_{ex} (1) definido en g/s.

$$\dot{m}_{ex} = \dot{m}_{in} + \dot{m}_f \quad (1)$$

Donde \dot{m}_{in} representa el flujo masico de aire y \dot{m}_f el flujo de combustible, los contaminantes se miden en base seca y se corrigen mediante la concentración húmeda $C_{wet,j}$ (2) expresada en unidades de volumen y $C_{dry,j}$ que representa la concentración en base seca además un factor de corrección nombrado como k_w (3), donde α representa una relación molar que describe al hidrógeno.

$$C_{wet,j} = k_{w,j} C_{dry,j} \quad (2)$$

$$k_w = \frac{1.008}{1+0.005\alpha(C_{CO_2}+C_{CO})} \quad (3)$$

Se dispone de las emisiones másicas instantáneas de cada uno de los contaminantes $\dot{m}_{j,i}$ (4), donde $\mu_{j,i}$ relaciona la densidad individual con la densidad conjunta de los componentes.

$$\dot{m}_{j,i} = c_{j,i} \mu_{j,i} \dot{m}_{ex,i} 10^{-3} \quad (4)$$

Seguidamente, la emisión de cada contaminante m_j (g) es la sumatoria de $\dot{m}_{j,i}$ (4) con respecto al tiempo como se describe en la ecuación (5) donde n representa el número de muestras y Δt el tiempo de muestreo.

$$m_j = \sum_{i=1}^n \dot{m}_{j,i} \Delta t \quad (5)$$

Finalmente, los factores de emisión $F_{j,k}$ de cada contaminante (g/km) acorde a la ruta RDE se define mediante la ecuación 6.

$$F_{j,k} = \frac{m_{j,k}}{s_k} \quad (6)$$

2.2. Materiales

A continuación, se presentan los materiales y equipos relacionados con la obtención de las y emisiones contaminantes.

2.2.1. Vehículo de pruebas

El vehículo seleccionado cumple con los estándares especificados por la normativa INEN 2656 [16], este es un SUV Kia Sportage del año 2018. Este vehículo proporciona un contexto relevante y representativo para la evaluación de los parámetros de rendimiento del motor y la eficiencia del sistema de inyección de combustible en condiciones reales de funcionamiento. Está equipado con una transmisión de seis velocidades y un motor 2.0 DOHC.

2.2.2. PEMS

El sistema de medición de emisiones portátil PEMS, es un equipo Kane Plus que mantiene un retraso de 100 ms en la medición, este valor es óptimo para lograr la mayor exactitud y fiabilidad en los resultados, así como para asegurar datos de forma eficaz y precisa durante todo el proceso de ensayo. Las mediciones de los gases dióxido de carbono (CO₂), monóxido de carbono (CO), oxígeno (O₂), hidrocarburos (HC) y los valores de óxidos nitrosos (NO_x) se obtienen de forma puntual [11- 17]. La eficacia y la consistencia del instrumento aseguran la calidad y precisión de los resultados presentados en los diversos rangos, tal como se detalla en la Tabla 1.

Tabla 1. Señal de sensores y unidades.

Parámetros	Rango	Unidad
CO ₂	0 – 18	[%]
CO	0 – 4.5	[%]
O ₂	0 – 22	[%]
HC	0 – 365	[ppm]
NO _x	0 – 360	[ppm]
Lambda	0 – 1.4	-

2.2.3. Registrador de datos OBD

El dispositivo Freematics ONE+ instalado en el puerto OBD II del vehículo captura las señales con una frecuencia de 10 Hz. A través de la programación incorporada en el mismo, se registran en tiempo real los datos provenientes de los sensores del vehículo [20], como a su vez se capturan los datos de longitud, latitud y altitud que describen la ruta presentada en la Figura 2.

Tabla 2. Señal de sensores y unidades.

Parámetros	Rango	Unidad
A _x	-7.1 – 3.4	[g/s]
MAP	16 – 74	[kPa]
RPM	0 – 3183	[rpm]
TPS	10 – 87	[%]
VSS	0 – 93	[Km/h]
O ₂	0 – 0.89	[V]
STFT	-12 – 29	[%]
LTFT	0 – 42	[%]

2.3. Importancia de variables

En la base de datos que contienen las señales PID's y emisiones contaminantes están presentes 24 tipos de datos recolectadas de los cuales se necesitan considerar las variables más influyentes en las diferentes condiciones de operación planteadas [21]. Dado que cada variable tiene un valor de relevancia, se implementa el algoritmo de aprendizaje automático Random Forest, ampliamente utilizado en tareas de regresión y clasificación. Al ser entrenado con varias las variables previamente obtenidas, este algoritmo se encarga de tomar una decisión final con respecto a al valor de importancia con un coeficiente de determinación $R^2 = 0.9486$ donde se especifican las 4 variables más determinantes: Aceleración (a_x), Dióxido de Carbono (CO₂), Óxidos Nitrosos (NO_x) y la velocidad del vehículo (VSS) en los diferentes escenarios establecidos, como se muestra en la Figura 3 y en la Tabla 3.

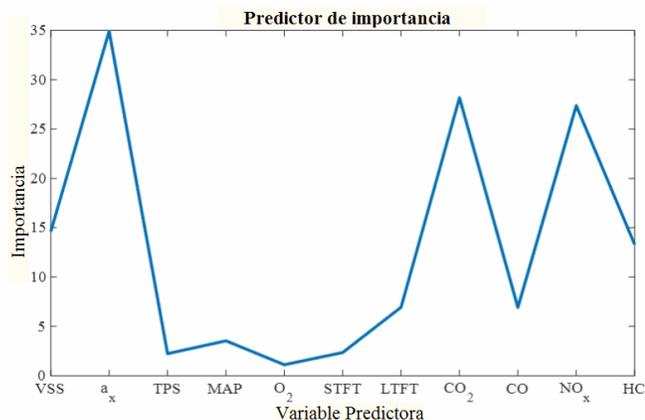


Fig. 3. Predictor de importancia de variables de Random Forest.

Tabla 3. Datos relevantes según Random Forest.

Parámetros	Símbolo	Valor de importancia
Aceleración	a_x	34.92
Dióxido de Carbono	CO ₂	28.18
Óxidos Nitrosos	NO _x	27.37
Velocidad del Vehículo	VSS	14.60
Hidrocarburos no combustionados	HC	13.29
Ajuste de largo plazo de combustible	LTFT	6.94
Monóxido de Carbono	CO	6.91
Presión Absoluta en la Admisión	MAP	3.53
Ajuste de corto plazo de combustible	STFT	2.35
Posición de la Mariposa de aceleración	TPS	2.23
Sensor de Oxígeno	O ₂	1.11

De esta forma Random Forest señala que las variables que más experimentan mayor número de variaciones según el modo de operación son 2 señales PID's y 2 emisiones contaminantes que comprenden un valor de importancia que comprende desde 14.29 hasta un 34.92, de este modo, cada variable tiene un comportamiento representativo que manifiesta una diferencia en la densidad de los datos en los diferentes modos de operación, como se observa en la Figura 4.

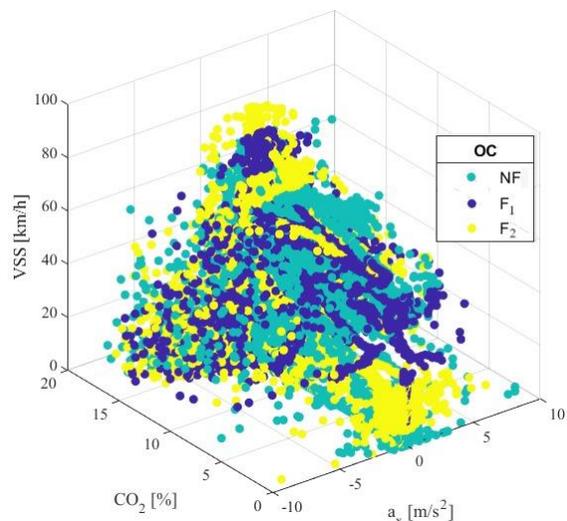


Fig. 4. Densidad de datos según Random Forest.

2.4. Entrenamiento de aprendizaje automático

El entrenamiento de los algoritmos de aprendizaje automático se da mediante matrices etiquetadas que representan las distintas pruebas realizadas de NF, F1 y F2, de esta forma, cada uno de los métodos de aprendizaje automático utilizan la matriz Totaltest (7) donde se incluyen las 4 variables más relevantes que se presentan en la Tabla 4.

$$Totaltest = [NF, F1, F2] \quad (7)$$

De esta manera, la creación de las bases de entrenamiento utiliza 4 modelos de aprendizaje supervisado [22] que presentan diferentes estructuras en su entrenamiento, en cuanto a las DecisionTrees el número de Split máximo es de 100 y un índice de diversidad de Gini, el modelo de K- Nearest Neighbors presenta de igual forma un total de 100 vecinos con una distancia Euclidiana entre dos puntos, el Self – Organizing Map aplica una función Kernely, y utiliza una codificación de multicategoría de Uno a Uno en un nivel de restricción de caja de 1, por último la Neural Network dispone de una máximo de 1000 iteraciones con 10 neuronas en la capa oculta. En cuanto a las características de cada una la variación en el tiempo de entrenamiento y el porcentaje de validación se presentan en la Tabla 4.

Tabla 4. Porcentajes de validación de cada modelo de aprendizaje automático

Parámetros	Tiempo de entrenamiento	Porcentaje de validación
Arboles de decisión	15.433 seg	84.93 %
SVM	4167.8 seg	60.57 %
KNN	9.0183 seg	88.50 %
Redes Neuronales	133.62 seg	82.43 %

Considerando los resultados mostrados en la Tabla 4 se descarta el uso del aprendizaje automático por SVM tanto por el porcentaje de validación como por ser el método que mayor tiempo requiere, en cuanto a los demás, aunque se presentan valores de validación parecidos entre sí, el modelo de aprendizaje automático KNN con 88.50% de validación y el menor tiempo en el entrenamiento es la mejor opción que se debe aplicar al modelo de predicción de fallas.

3. Resultados y Discusión

A continuación, después del análisis de datos y el entrenamiento de los modelos de aprendizaje automático, es momento de analizar los resultados y principalmente la efectividad que proporciona el modelo en cuanto a la detección de fallos.

3.1. K – Nearest Neighbors

En conformidad con los resultados de validación de cada uno de los modelos de aprendizaje automático, se considera a los K – Nearest Neighbors como el método ideal, ahora, en la Figura 5 está presente el porcentaje adquirido por cada una de las condiciones de operación en relación con los Valores Predictivos Positivos (PPV) y las Tasas de Descubrimiento Falsas (FDR).

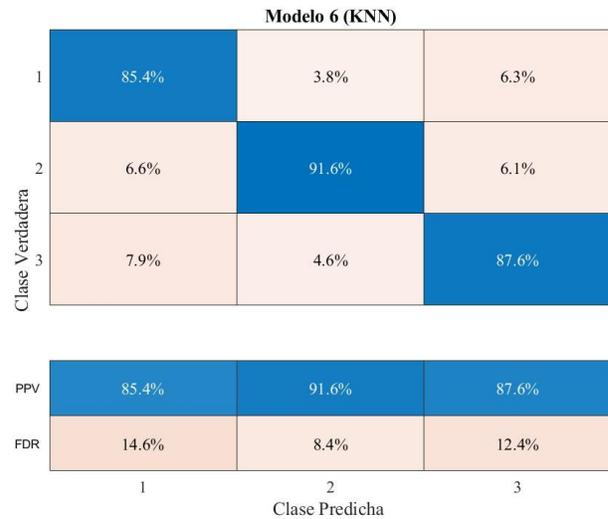


Fig. 5. Porcentaje de validación de K – Nearest Neighbor.

En cuanto a la preparación de esta inteligencia artificial se especifican 4 entradas características determinadas por la relevancia según Random Forest y 3 opciones de salida que determinan si existen o no fallas en el motor. En la figura 6 y 7 se representa el modelo de predicción con respecto a las variables de Aceleración, Óxidos Nitrosos (NO_x) y Dióxido de Carbono (CO₂).

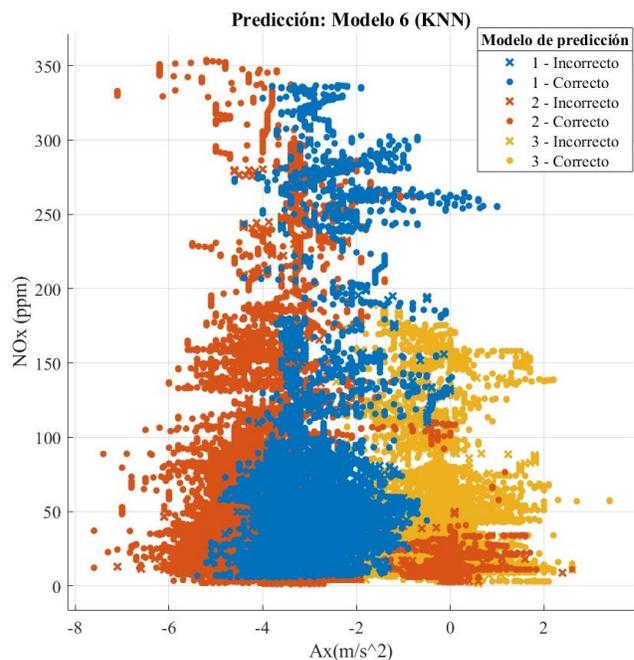


Fig. 6. Predicciones de K – Nearest Neighbors con aceleración y NO_x

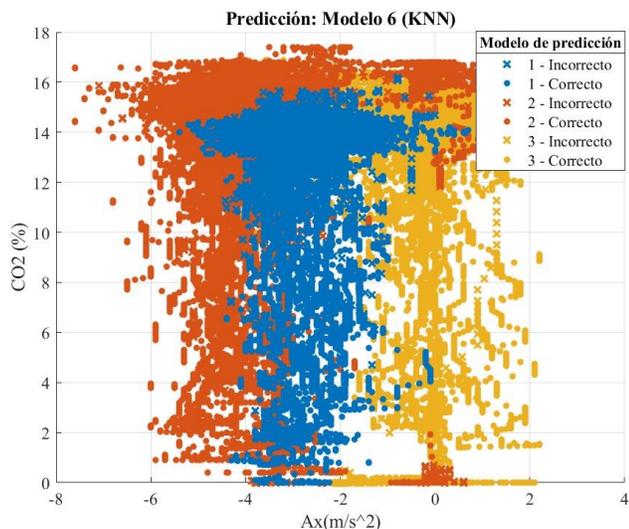


Fig. 7. Predicciones de K – Nearest Neighbors con aceleración y CO₂.

En cuanto al rendimiento del modelo de detección de fallas, la curva de ROC obtiene un valor de área bajo de curva (AUC) de 0.9769, como se muestra en la figura 8.

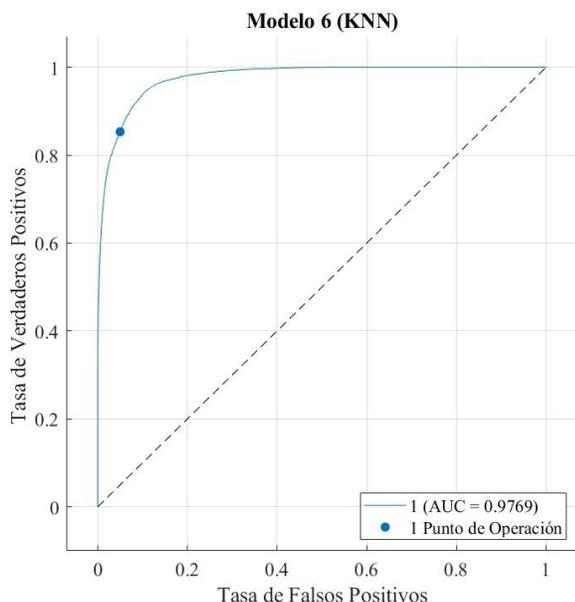


Fig. 8. Curva de ROC del modelo KNN.

3.2. Validación del modelo

Los datos obtenidos en las pruebas de ruta RDE y aleatorias se utilizan como insumos para el modelo de detección de fallas [7-23]. Estos datos, que incluyen los tres comportamientos NF, F1 y

F2, se agrupan en conjuntos de 100 y se reorganizan aleatoriamente en una nueva matriz que en efecto son datos de ensayo para su clasificación mediante el uso modelo de detección de fallas, como se muestra en la Tabla 5.

Tabla 5. Modelo de validación.

Condición	Cantidad Real de Datos	Predicción de Datos por KNN	Error
NC	23567 units	23508 units	59 units
F1	35635 units	35172 units	436 units
F2	33113 units	33635 units	522 units

De acuerdo con los resultados de la Tabla 5, el máximo número de error son 522 datos fuera su condición original, es decir, que existe un porcentaje de error máximo de 1.37 % como se muestra en la Figura 9.

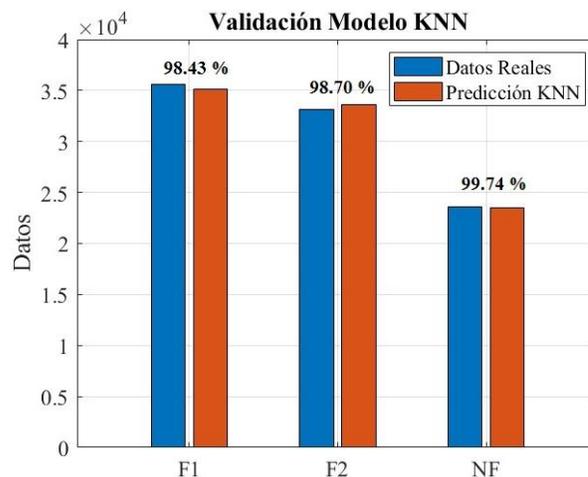


Fig 9. Validación del Modelo KNN con datos aleatorios.

3.3. Modelo de predicción de fallos

El modelo de predicción de fallas en el sistema de inyección ha demostrado un rendimiento sobresaliente en términos de clasificación de las diferentes condiciones de funcionamiento del motor, K – Nearest Neighbors como modelo de aprendizaje automático utiliza las variables Ax, CO₂, NO_x y VSS para determinar las condiciones de NF, F1 y F2, mediante una prueba de datos aleatorios los resultados se presentan en la Tabla 6.

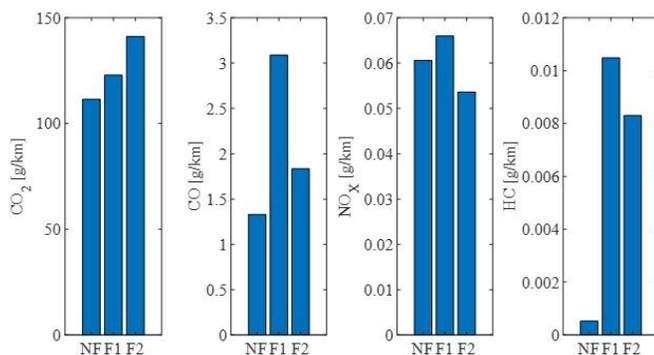
Tabla 6. Porcentaje de error del modelo KNN

	NC	F1	F2
Error	0.26%	1.30%	1.36%

3.4. Factores de emisiones

Las variables se derivan de un análisis realizado sobre la base de datos de emisiones, empleando la metodología RDE donde las emisiones tienden a incrementarse en situaciones de funcionamiento anormal del motor en comparación con su estado NF, como se muestra en la Figura 10. Específicamente, se observa un aumento proporcional del dióxido de carbono (CO_2), destacándose la presencia de F2 como más relevante, lo cual sugiere una deficiencia en el proceso de combustión.

De manera similar, el monóxido de carbono exhibe un incremento gradual, indicativo de una combustión incompleta. La respuesta de los inyectores ante la insuficiencia de ingreso de aire consiste en enriquecer la mezcla, lo que resulta en un aumento de la presión y la temperatura, y en valores más representativos de NO_x durante el modo de fallo F1 [24]. No obstante, los hidrocarburos muestran un índice elevado de contaminación, evidenciando que el combustible no cumple adecuadamente su ciclo de trabajo al salir del motor en una condición similar a la de ingreso. La anticipación de estas fallas puede propiciar una notable reducción de los gases contaminantes, contribuyendo así a un entorno menos perjudicial desde el punto de vista ambiental y de salud pública.

**Fig. 10.** Factores de emisión.**Tabla 7.** Factores de emisión

Emisiones	Unidad	Condición de Operación		
		NF	F1	F2
CO_2	g/kg	111.44	122.84	141.13
CO	g/kg	1.32	3.08	1.83
NO_x	g/kg	0.0606	0.0660	0.0536
Hc	g/kg	0.0005	0.0105	0.0083

4. Conclusiones

La investigación propone un modelo para la detección de fallas en el sistema de inyección, limitado a dos averías F1 y F2, utilizando aprendizaje automático combinado con señales PID's y las emisiones contaminantes. Este enfoque demuestra un alto índice de efectividad en la predicción de las condiciones propuestas, como se muestra en la Figura 9 y la Tabla 6, donde se detallan los porcentajes de error.

Las variables más influyentes en las condiciones de operación, estimadas mediante Random Forest, revelan que la aceleración es el factor principal con un rango de importancia de 34.92, seguida del CO_2 con 28.18, NO_x con 27.37 y la velocidad del vehículo con 14.60. Estos resultados indican que mayores aceleraciones conducen a velocidades más altas y a una mayor producción de emisiones contaminantes. Además, al enriquecer drásticamente la mezcla inyectando más combustible para compensar las fallas, se incrementan la presión y la temperatura, lo que a su vez influye en la oxidación del nitrógeno.

En el análisis de los factores de emisión destaca como cada una de las fallas tienen un impacto significativo en todas las emisiones, en caso de CO_2 se observa un incremento lineal siendo F2 el factor más severo, por otro lado el CO aumenta notablemente con F1 debido a la falta de aire y las correcciones de combustible, el mismo efecto representa el aumento de las emisiones de NO_x en este escenario, en cuanto a HC se observa un incremento drástico en las dos fallas provocadas con respecto al estado NF, destacando que anticipar estas fallas contribuye a prevenir los elevados índices de contaminación.

Respecto al aprendizaje automático es posible generar un modelo entrenado y supervisado utilizando las variables más relevantes. Entre las técnicas evaluadas, el uso de K-Nearest Neighbors se destaca como la mejor opción, alcanzando un porcentaje de validación del 88.50% y un valor de la curva ROC con valor bajo la curva de 0.9769. En comparación, otras herramientas de aprendizaje automático, como los Árboles de Decisión, las Redes Neuronales y el Self-Organizing Map, mostraron porcentajes de validación más bajos, con 84.93%, 82.43% y 60.57% respectivamente.

Este modelo predictivo a pesar de generar buenos resultados en la clasificación de fallas presenta comportamientos no esperados en los datos recolectados, las señales de los sensores de temperatura ECT (Temperatura del refrigerante del motor) e IAT (Temperatura del aire de admisión) contienen gran fluctuación de datos y según un análisis de correlación se imponían como las variables más representativas de los fallos provocados, cuando en realidad este comportamiento es resultado del funcionamiento estándar del motor en trayectos de alta velocidad. El modelo predictivo esta entrenado con variables de velocidad y aceleración, misma que en un inicio no eran punto de análisis pero que resultaron ser determinantes después de la estimación de relevancia por medio de Random Forest.

Finalmente, el modelo de predicción está construido en base a 2 fallas básicas que ocurren en el sistema de inyección, es decir, la búsqueda para abarcar todos los posibles escenarios de falla que se producen en el motor, es mediante la introducción de averías de mayor gravedad y complejidad, resultando en una gran diversidad de datos para entrenar el modelo de predicción, como tal, este es el punto de partida para los diagnósticos anticipados de fallas mediante técnicas de aprendizaje automático.

Referencias

- [1] Bermeo Naula, P. Molina Campoverde, and N. R. Campoverde, "Methodological proposal for estimating polluting emissions: Case of Cuenca, Ecuador," in IOP Conference Series: Earth and Environmental Science, Institute of Physics, 2023. doi: 10.1088/1755-1315/1141/1/012003.
- [2] S. Muñoz Pérez, J. Salcedo Reátegui, and A. Sotomayor Mendoza, "Contaminación ambiental producida por el tránsito vehicular y sus efectos en la salud humana: revisión de literatura," INVENTUM, vol. 16, no. 30, pp. 20–30, Apr. 2021, doi: 10.26620/uniminuto.inventum.16.30.2021.20-30.
- [3] Z. Kan, L. Tang, M. P. Kwan, and X. Zhang, "Estimating vehicle fuel consumption and emissions using GPS big data," Int J Environ Res Public Health, vol. 15, no. 4, Apr. 2018, doi: 10.3390/ijerph15040566.
- [4] N. Dhote et al., "Monitoring Vehicular Pollution By Using Embedded System." [Online]. Available: www.ijtrd.com.
- [5] O. Rybitskyi, V. Golian, N. Golian, Z. Dudar, O. Kalynychenko, and D. Nikitin, "Using obd-2 technology for vehicle diagnostic and using it in the information system," bulletin of national technical university "KhPI". Series: System Analysis, Control and Information Technologies, no. 1 (9), pp. 97–103, Jul. 2023, doi: 10.20998/2079-0023.2023.01.15.
- [6] M. A. K. Niazi et al., "Development of an On-Board Diagnostic (OBD) kit for troubleshooting of compliant vehicles," in ICET 2013 - 2013 IEEE 9th International Conference on Emerging Technologies, IEEE Computer Society, 2013. doi: 10.1109/ICET.2013.6743551.

- [7] N. Rivera, E. Jiménez, and J. Cárdenas, “Parametric Model for Estimating Pollutant Emissions in M1 Otto Cycle Vehicles with OBD-II,” in *ECTM 2023 - 2023 IEEE 7th Ecuador Technical Chapters Meeting, Institute of Electrical and Electronics Engineers Inc.*, 2023. doi: 10.1109/ETCM58927.2023.10308992.
- [8] Ingeniería Mecánica Automotriz, C. DE, Landin Carlos Cesar Rivera Caravajo Juan Andrés Tutor, C., & Néstor Diego Rivera Campoverde, I. (n.d.). Diagnóstico de fallas de un motor de encendido provocado Corsa Evolution 1.4L mediante el análisis de las señales de los sensores MAP, TPS, VSS, utilizando herramientas de aprendizaje y clasificación.
- [9] Ernesto, I., Llanzhi, F., Santiago, H., Inga, I., Néstor, I., & Rivera Campoverde, D. (n.d.). Análisis de modos de falla del sistema de inyección electrónica de combustible multec Delphi.
- [10] Vergara, M., Ramos, L., Rivera-Campoverde, N. D., & Rivas-Echeverria, F. (2023). EngineFaultDB: A Novel Dataset for Automotive Engine Fault Classification and Baseline Results. *IEEE Access*, 11, 126155–126171. <https://doi.org/10.1109/ACCESS.2023.3331316>.
- [11] N. Rivera-Campoverde, J. M. Sanz, and B. Arenas-Ramirez, “Low-Cost Model for the Estimation of Pollutant Emissions Based on GPS and Machine Learning,” in *Proceedings of the XV Ibero-American Congress of Mechanical Engineering*, Springer International Publishing, 2023, pp. 182–188. doi: 10.1007/978-3-031-38563-6_27.
- [12] Institute of Electrical and Electronics Engineers, 2019 IEEE 90th Vehicular Technology Conference (VTC2019-Fall): proceedings: Honolulu, Hawaii, USA, 22-25 September 2019.
- [13] E. A. Llanes-Cedeño, Y. Guardia-Puebla, A. de la Rosa-Andino, S. Cevallos-Carvajal, and J. C. Rocha-Hoyos, “Detection of Faults in Combustion Engines Through Indicators of Temperature and Injection Pressure,” *Ingenius*, vol. 2019, no. 22, pp. 38–46, Jul. 2019, doi: 10.17163/ings.n22.2019.04
- [14] R. kumar and A. Jain, “Driving behavior analysis and classification by vehicle OBD data using machine learning,” *Journal of Supercomputing*, vol. 79, no. 16, pp. 18800–18819, Nov. 2023, doi: 10.1007/s11227-023-05364-3.
- [15] N. D. Rivera-Campoverde, B. Arenas-Ramírez, J. L. Muñoz Sanz, and E. Jiménez, “GPS Data and Machine Learning Tools, a Practical and Cost-Effective Combination for Estimating Light Vehicle Emissions,” *Sensors*, vol. 24, no. 7, Apr. 2024, doi: 10.3390/s24072304.
- [16] “NTE INEN 2656 CLASIFICACIÓN VEHICULAR VEHICLE CLASIFICATION 31 Páginas,” 2016.
- [17] P. A. Molina Campoverde, “Estimation of Fuel Consumption through PID Signals Using the Real Emissions Cycle in the City of Quito, Ecuador,” *Sustainability (Switzerland)*, vol. 15, no. 16, Aug. 2023, doi: 10.3390/su151612474
- [18] Unión Europea, “REGLAMENTO (UE) 2017/1154,” pp. 708–732, Jun. 2017.
- [19] Rivera Campoverde, N. D., Chica Segovia, J. F., Zambrano, I., & Garcia Garcia, C. L. (2017). Estudio del Comportamiento de un Motor Ciclo Otto de Inyección Electrónica Respecto de la Estequiometría de la Mezcla y del Adelanto al Encendido para la Ciudad de Cuenca. *Revista Politécnica*, 40(40), 1-9.

- [20] Molina Campoverde, P. A., Rivera Campoverde, N. D., Novillo Quirola, G. P., & Bermeo Naula, A. K. (2021). Characterization of Braking and Clutching Events of a Vehicle Through OBD II Signals. En M. Botto-Tobar, W. Zamora, J. Larrea Plúa, J. Bazarro Roldan, & A. Santamaría Philco (Eds.), *Systems and Information Sciences - Proceedings of ICCIS 2020* (pp. 134-143). (Advances in Intelligent Systems and Computing; Vol. 1273 AISC). Springer Science and Business Media Deutschland GmbH. https://doi.org/10.1007/978-3-030-59194-6_12
- [21] F. Rosero Obando, X. Rosero, and Z. Mera, “Developing Fuel Efficiency and CO2 Emission Maps of a Vehicle Engine Based on the On-Board Diagnostic (OBD) Approach,” *Enfoque UTE*, vol. 15, no. 1, pp. 7–15, Jan. 2024, doi: 10.29019/enfoqueute.1002.
- [22] P. Czech, G. Wojnar, R. Burdzik, Ł. Konieczny, and J. Warczek, “268. Application of the discrete wavelet transform and probabilistic neural networks in IC engine fault diagnostics,” vol. 16, no. 4, 2014.
- [23] Y. Shatnawi and M. Al-Khassaweneh, “Fault diagnosis in internal combustion engines using extension neural network,” *IEEE Transactions on Industrial Electronics*, vol. 61, no. 3, pp. 1434–1443, 2014, doi: 10.1109/TIE.2013.2261033.
- [24] I. Zibani, K. Tsamaase, E. Matlotse, and I. Ngebani, “An Integrated Engine Control Unit for a Spark Ignited Internal Combustion Engine,” *European Journal of Electrical Engineering and Computer Science*, vol. 7, no. 6, pp. 1–5, Nov. 2023, doi: 10.24018/ejece.2023.7.6.574.