

**PONTIFICIA UNIVERSIDAD
CATÓLICA DEL PERÚ**

Escuela de Posgrado



ESTIMACIÓN DEL CONSUMO DE COMBUSTIBLE Y EMISIONES DE CO₂
DE UN MOTOR A GASOLINA DE 130 HP MEDIANTE TÉCNICAS DE
MACHINE LEARNING

Tesis para obtener el grado académico de Maestro en Energía que
presenta:

José Huancapaza Machuca

Asesor:

Dr. Julio César Cuisano Egúsquiza

Lima, 2023


Informe de Similitud

Yo, Julio César Cuisano Egúsqiiza, docente de la Escuela de Posgrado de la Pontificia Universidad Católica del Perú, asesor de la tesis “ESTIMACIÓN DEL CONSUMO DE COMBUSTIBLE Y EMISIONES DE CO₂ DE UN MOTOR A GASOLINA DE 130 HP MEDIANTE TÉCNICAS DE MACHINE LEARNING”, del autor José Huancapaza Machuca dejo constancia de lo siguiente:

- El mencionado documento tiene un índice de puntuación de similitud de 14%. Así lo consigna el reporte de similitud emitido por el software *Turnitin* el 29/10/2023.
- He revisado con detalle dicho reporte y la Tesis, y no se advierte indicios de plagio.
- Las citas a otros autores y sus respectivas referencias cumplen con las pautas académicas.

Lugar y fecha:

San Miguel, 29 de octubre de 2023.

Cuisano Egúsqiiza, Julio César	
DNI: 10744493	
ORCID:0000-0002-2175-3656	

DEDICATORIA

A mi mamá, hermanos y sobrinos.



AGRADECIMIENTOS

Al Doctor Julio César Cuisano Egúsqiza, mi asesor por su gran apoyo para la realización de este trabajo.

A mi mamá, hermanos y sobrinos, por su paciencia.



RESUMEN

En el presente trabajo se estima el consumo de gasolina y las emisiones de CO₂ en un motor vehicular de 130 HP, instalado en un banco de pruebas, usando técnicas de aprendizaje automático (Machine Learning). Para obtener datos de los parámetros de funcionamiento del motor, se realizaron pruebas en condiciones estacionarias de carga (torque) y régimen de giro del cigüeñal; se registraron las lecturas de sensores originalmente instalados en el motor, mediante un scanner conectado al puerto OBD2 del módulo de control electrónico del motor. Además, se instalaron equipos de laboratorio para registrar otras variables necesarias para el estudio. Con los datos disponibles, se utilizaron tres técnicas de Machine Learning: Regresión Múltiple, Máquina de Soporte Vectorial y Redes Neuronales.

En la aplicación de los modelos se utilizaron datos en grupos, separados de la siguiente forma: 90% para el desarrollo de los modelos y 10 % para la prueba de los modelos. Adicionalmente, para los modelos de Máquina de Soporte Vectorial y de Redes Neuronales se realizó otra partición de los datos: 75% para entrenamiento, 15% para validación, y 15% para el test. Durante el proceso se evaluaron los datos sin estandarización y, posteriormente, estandarizados en el rango de 0 a 1; este último paso buscó asegurar la convergencia del modelo.

Las variables estudiadas fueron las siguientes: i) 5 predictoras o variables independientes (presión absoluta en el colector de admisión, temperatura del aire en el colector de admisión, régimen de giro, flujo másico de aire de

admisión al motor y el torque efectivo); ii) 2 variables objetivo o dependientes (emisiones de CO₂ y consumo de gasolina).

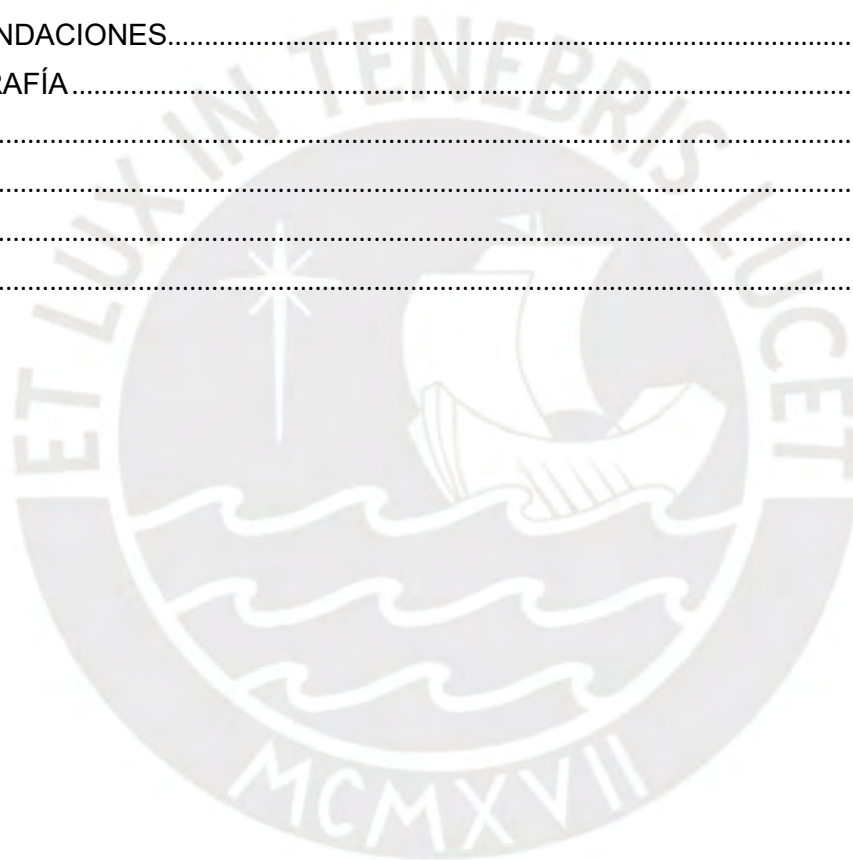
Los resultados del presente trabajo de tesis muestran que el mejor método, y con menos intervención, es el de Redes Neuronales. Para la estimación del flujo másico instantáneo del CO₂ se obtuvo un error máximo de 7.85%, siendo que el error obtenido para el 75% de los resultados corresponde a 0.10%. Para la estimación del consumo másico de gasolina, se obtuvo un error máximo de 9.72%, pero, en este caso, el 75% de los resultados tienen un error de 0.67%.



CONTENIDO

DEDICATORIA.....	ii
AGRADECIMIENTOS.....	iii
RESUMEN.....	iv
CONTENIDO.....	vi
INTRODUCCIÓN.....	1
Definición del problema.....	3
Propuesta de Solución.....	3
Justificación.....	4
Objetivos.....	5
Objetivo general.....	5
Objetivo específico.....	5
CAPÍTULO 1.....	6
FUNDAMENTO TEÓRICO.....	6
1.1. Scanner OBD.....	6
1.2. Machine Learning.....	8
1.2.1. Regresión Lineal Múltiple.....	8
1.2.2. Máquina de Soporte Vectorial.....	12
1.2.3. Redes neuronales.....	14
1.3. Fórmulas para el análisis.....	19
CAPÍTULO 2.....	22
ESTADO DEL ARTE.....	22
CAPÍTULO 3.....	35
DESARROLLO.....	35
3.1. Equipos empleados:.....	36
3.2. Etapas de la experimentación.....	45
Primera etapa de la experimentación:.....	45
3.3. Implementación de los modelos de Machine Learning.....	57
3.4. Condición de repetibilidad.....	58
3.5. Preprocesamiento de datos.....	58
3.6. División de datos.....	62
3.7. Estandarizar datos.....	65
3.8. Procesamiento de datos y generación de modelos.....	65
3.8.1. Regresión Lineal Múltiple.....	66
3.8.2. Máquina de soporte Vectorial (SVM).....	69
3.8.3. Redes Neuronales.....	71
CAPÍTULO 4.....	74

RESULTADOS Y ANÁLISIS.....	74
4.1. Modelos para ajustar emisiones de CO ₂	74
4.1.1. Regresión Lineal Múltiple	75
4.1.2. Máquina de Soporte Vectorial	83
4.1.3. Redes neuronales	86
4.2. Modelos para ajustar el consumo de combustible	94
4.2.1. Regresión Lineal Múltiple	94
4.2.2. Máquina de Soporte Vectorial	99
4.2.3. Redes neuronales	104
4.3. Resumen de resultados para los modelos	111
CONCLUSIONES.....	114
RECOMENDACIONES.....	118
BIBLIOGRAFÍA.....	119
Anexo A.....	122
Anexo B.....	123
Anexo C	124
Anexo D	125



ÍNDICE DE FIGURAS

FIGURA 1 - INTERFACE ELM327 [3].	7
FIGURA 2 - DESCRIPCIÓN DE TERMINALES OBD2.	7
FIGURA 3 - COOK'SDISTANCE-TOMADO DE MATWORKS [5]	11
FIGURA 4 - LEVERAGE-TOMADO DE MATWORKS [10]	12
FIGURA 5 - DATOS SEPARADOS POR UNA LÍNEA [11].	13
FIGURA 6 - DATOS SEPARADOS POR UN PLANO [11].	13
FIGURA 7 - APRENDIZAJE SUPERVISADO. FUENTE: REDES NEURONALES ARTIFICIALES [12].	16
FIGURA 8 - PERCEPTRON NETWORK, TOMADO DE NND NETWORK DESING [15]	18
FIGURA 9 - RED NEURONAL, ELABORACIÓN PROPIA.	19
FIGURA 10 - EMISIONES DE CO ₂ REALES Y PREDICHO. TOMADO DE REFERENCIA [17].(DOI: 10.3390/ATMOS13091466).	24
FIGURA 11 - EMISIONES DE CO ₂ , REAL Y PREDICCIÓN. TOMADO DEL MODELO DE EMISIONES PUBLICADO POR MADZIEL Y COLEGAS [18].(DOI: 10.3390/EN15010142).	26
FIGURA 12 - AJUSTE DE CORRELACIÓN FASE TEST BBDT TOMADO DE REFERENCIA [23].(DOI: 10.4172/2573-458X.1000106).	31
FIGURA 13 - MÓDULO DE CONTROL DEL DINAMÓMETRO.	38
FIGURA 14 - MÓDULO DE CONTROL DEL MOTOR HYUNDAI 1600CC.	38
FIGURA 15 - DINAMÓMETRO CONECTADO AL MOTOR CON UN EJE CARDAN	39
FIGURA 16 - MEDIDOR DE AIRE DE FLUJO LAMINAR.	40
FIGURA 17 - VENTILADOR AXIAL	40
FIGURA 18 - TRANSMISOR Y MEDIDOR DE FLUJO TIPO CORIOLIS.	41
FIGURA 19 - EJX110A TRANSMISOR DE PRESIÓN DIFERENCIAL.	42
FIGURA 20 - TACÓMETRO MONARCH, MODELO ACT-3X	42
FIGURA 21 - REGISTRADOR YOKOGAWA.	43
FIGURA 22 - ANALIZADOR DE GASES DE ESCAPE VEHICULAR, MARCA INFRARED INDUSTRIES, MODELO FGA4500.	44
FIGURA 23 - DIAGRAMA DE FLUJO DE PASOS A SEGUIR EN PRUEBAS DE VERIFICACIÓN.	45
FIGURA 24 - ACCESORIOS DE CONEXIÓN Y MEDIDOR DEL FLUJO DE AIRE DE ADMISIÓN AL MOTOR.	46

<i>FIGURA 25 - VISTA GENERAL DEL MEDIDOR DE FLUJO DE AIRE Y BANCO DE PRUEBA DEL MOTOR.</i>	<i>46</i>
<i>FIGURA 26 - MEDIDOR DE PRESIÓN DIFERENCIAL DWYER 477</i>	<i>47</i>
<i>FIGURA 27 - DETALLE DE CONEXIÓN ENTRE MOTOR Y DINAMÓMETRO MEDIANTE EJE CARDAN.</i>	<i>47</i>
<i>FIGURA 28 - DINAMÓMETRO EN EL RANGO DE INFERIOR CON UN MÁXIMO DE 150 NM.....</i>	<i>48</i>
<i>FIGURA 29 - MATERIALES USADOS Y TRANSMISOR CONECTADO A UN MEDIDOR CORIOLIS (LECTURA EN KG/H).....</i>	<i>48</i>
<i>FIGURA 30 - VISOR DE RPM Y CONTROL DE TEMPERATURA ASOCIADO EL DINAMÓMETRO.</i>	<i>49</i>
<i>FIGURA 31 - VENTILADOR PARA REGULAR LA TEMPERATURA DEL MOTOR EN BANCO DE PRUEBAS.</i>	<i>49</i>
<i>FIGURA 32 - DISPOSITIVO OBDII CONECTADO A TELEFONO CELULAR VÍA APLICATIVO PARA CAPTURA DE DATOS EN TIEMPO REAL.</i>	<i>50</i>
<i>FIGURA 33 - MEDICIÓN DE FUERZA PARA MOVER PEDAL DE ACELERADOR.....</i>	<i>53</i>
<i>FIGURA 34 - DISEÑO DE MESA PARA INSTALACIÓN DE ACTUADOR LINEAL "V-SLOT".....</i>	<i>54</i>
<i>FIGURA 35 - PROCESO DE FABRICACIÓN Y USO DEL SOPORTE DEL ACTUADOR Y CONTROLADOR DEL PEDAL.....</i>	<i>54</i>
<i>FIGURA 36 - SOPORTE INSTALADO EN EL PANEL DE CONTROL, CON ACTUADOR LINEAL INSTALADO (INCLUIDO MOTOR NEMA 23).....</i>	<i>55</i>
<i>FIGURA 37 - INSTALACIÓN DE REGISTRADOR Y CAMBIO DE MEDIDOR DIFERENCIAL DE PRESIÓN DWYER 477 POR YOCOGAWA MODELO E110A.</i>	<i>56</i>
<i>FIGURA 38 - CAPTURA DE PANTALLA DEL REGISTRADOR YOCOGAWA DURANTE SU CONFIGURACIÓN POR TÉCNICO DE LABORATORIO.....</i>	<i>56</i>
<i>FIGURA 39 - DISPERSIÓN DE DIFERENTES VARIABLES VS CO₂.....</i>	<i>60</i>
<i>FIGURA 40 - DISPERSIÓN DE DIFERENTES VARIABLES VS CONSUMO DE GASOLINA. ELABORACIÓN PROPIA.</i>	<i>61</i>
<i>FIGURA 41 - PARTICIÓN DE DATOS: ENTRENAMIENTO (IZQUIERDA), TEST (DERECHA).</i>	<i>63</i>
<i>FIGURA 42 - PARTICIÓN DE DATOS: ENTRENAMIENTO (IZQUIERDA), TEST (DERECHA).</i>	<i>64</i>
<i>FIGURA 43 - DIAGRAMA DE FLUJO PARA ALGORITMO DE REGRESIÓN LINEAL MÚLTIPLE.....</i>	<i>68</i>
<i>FIGURA 44 - DIAGRAMA DE FLUJO PARA ALGORITMO DE MAQUINA DE SOPORTE VECTORIAL PARA REGRESIÓN.....</i>	<i>70</i>

<i>FIGURA 45 - DIAGRAMA DE FLUJO PARA ALGORITMO CON RED NEURONAL CON VARIACIÓN DE NEURONAS EN LA CAPA OCULTA.....</i>	<i>72</i>
<i>FIGURA 46 - DIAGRAMA DE FLUJO PARA CRITERIO DE PARADA DE RED NEURONAL.....</i>	<i>73</i>
<i>FIGURA 47 - DATOS OBSERVADOS Y ESTIMADOS PARA CO₂ CON DATOS DE ENTRENAMIENTO (MODELO 'QUADRATIC').</i>	<i>76</i>
<i>FIGURA 48 - DATOS OBSERVADOS Y ESTIMADOS PARA CO₂ CON DATOS DE TEST (MODELO 'QUADRATIC').</i>	<i>76</i>
<i>FIGURA 49 - COMPARACIÓN VALORES OBSERVADOS Y ESTIMADOS PARA EMISIÓN DE CO₂ PARA DATOS DE TEST.</i>	<i>77</i>
<i>FIGURA 50 - DIAGRAMA DE CAJAS DEL ERROR RELATIVO CON 115 OUTLIERS PARA AJUSTE DE EMISIONES DE CO₂ PARA DATOS DE TEST.</i>	<i>77</i>
<i>FIGURA 51 - COMPARACIÓN VALORES OBSERVADOS Y ESTIMADOS PARA EMISIÓN DE CO₂ SIN REGISTROS QUE TIENEN UNA DISTANCIA DE COOK > 3 VECES MEDIA DISTANCIA DE COOK.....</i>	<i>78</i>
<i>FIGURA 52 - DIAGRAMA DE CAJAS DEL ERROR RELATIVO CON 105 OUTLIERS PARA AJUSTE DE EMISIONES DE CO₂ PARA DATOS DE TEST.</i>	<i>79</i>
<i>FIGURA 53 - COMPARACIÓN VALORES OBSERVADOS Y ESTIMADOS PARA EMISIÓN DE CO₂ CON REGISTROS QUE TIENEN UNA DISTANCIA DE COOK < 5 MEDIA DISTANCIA DE COOK.</i>	<i>80</i>
<i>FIGURA 54 - DIAGRAMA DE CAJAS DEL ERROR RELATIVO CON 97 OUTLIERS PARA AJUSTE DE EMISIONES DE CO₂-COOK<=5 VECES MEDIA DISTANCIA DE COOK PARA DATOS DE TEST</i>	<i>81</i>
<i>FIGURA 55 - COMPARACIÓN VALORES OBSERVADOS Y ESTIMADOS PARA CO₂, STANDARDIZE, KERNEL POLINOMIAL DE ORDEN 2 (ENTRENAMIENTO) SVM.....</i>	<i>83</i>
<i>FIGURA 56 - COMPARACIÓN VALORES OBSERVADOS Y ESTIMADOS PARA CO₂, STANDARDIZE, KERNEL POLINOMIAL DE ORDEN 2 (TEST) SVM.</i>	<i>84</i>
<i>FIGURA 57 - VALORES OBSERVADOS VERSUS ESTIMADOS PARA EMISIONES DE CO₂- STANDARDIZE, KERNEL POLINOMIAL DE ORDEN 2 SVM.</i>	<i>84</i>
<i>FIGURA 58 - DIAGRAMA DE CAJAS PARA ERROR RELATIVO- STANDARIZE, KERNEL POLINOMIAL DE ORDEN 2 CON 293 OUTLIERS PARA DATOS DE TEST.....</i>	<i>85</i>
<i>FIGURA 59 - RMSE (ENTRENAMIENTO) DEL FLUJO DE CO₂ EN FUNCIÓN DE CANTIDAD DE NEURONAS PARA ESTIMADO DE EMISIÓN DE CO₂.....</i>	<i>87</i>

<i>FIGURA 60 - EVOLUCIÓN DE ERROR RELATIVO ESTIMADO DE CO₂ DATOS DE TEST - TABLAS 18,19 Y 20</i>	
.....	89
<i>FIGURA 61 - RMSE PARA 1 A 20 NEURONAS PARA ESTIMADO DE EMISIONES DE CO₂</i>	90
<i>FIGURA 62 - MSE PARA ENTRENAMIENTO, VALIDACIÓN Y TEST CON 13 NEURONAS PARA ESTIMADO DE EMISIÓN DE CO₂</i>	90
<i>FIGURA 63 - ENTRENAMIENTO DE LA RED CON 13 NEURONAS PARA PREDICCIÓN DE CO₂ CON REDES NEURONALES</i>	91
<i>FIGURA 64 - DIAGRAMAS DE CORRELACIÓN PARA PREDICCIÓN DE EMISIONES DE CO₂ REDES NEURONALES, PARA DATOS DE ENTRENAMIENTO (IZQUIERDA) Y TEST (DERECHA)</i>	91
<i>FIGURA 65 - VALORES OBSERVADOS VERSUS PREDICCIÓN PARA EMISIONES DE CO₂ CON 13 NEURONAS EN LA CAPA OCULTA</i>	92
<i>FIGURA 66 - DIAGRAMA DE CAJAS DE ERROR RELATIVO CON 100 OUTLIERS, RED CON 13 NEURONAS EN LA CAPA OCULTA</i>	92
<i>FIGURA 67 - DIAGRAMA DE CAJAS DE ERROR RELATIVO PARA TEST, CON 95 OUTLIERS PARA EMISIONES DE CO₂ USANDO DATOS FILTRADOS POR DISTANCIA DE COOK, RED NEURONAL CON 13 NEURONAS EN LA CAPA OCULTA</i>	93
<i>FIGURA 68- DATOS OBSERVADOS Y PRONÓSTICO PARA CONSUMO DE COMBUSTIBLE, DATOS DE ENTRENAMIENTO ("INTERACTION")</i>	95
<i>FIGURA 69 - DATOS OBSERVADOS Y PRONÓSTICO PARA CONSUMO DE COMBUSTIBLE, DATOS DE TEST (MODELO "INTERACTIONS")</i>	96
<i>FIGURA 70 - COMPARACIÓN VALORES OBSERVADOS Y ESTIMADO PARA CONSUMO DE COMBUSTIBLE</i>	96
<i>FIGURA 71 - VALORES OBSERVADOS Y ESTIMADOS PARA EL CONSUMO DE COMBUSTIBLE EN EL RANGO DE [300,400]</i>	97
<i>FIGURA 72 - DIAGRAMA DE CAJAS DEL ERROR RELATIVO CON 211 OUTLIERS PARA AJUSTE DE CONSUMO COMBUSTIBLE</i>	97
<i>FIGURA 73 - COMPARACIÓN VALORES OBSERVADOS Y AJUSTADO PARA EL CONSUMO DE COMBUSTIBLE, KERNEL POLINOMIAL DE ORDEN2 (ENTRENAMIENTO) PARA SVM</i>	101
<i>FIGURA 74 - COMPARACIÓN VALORES OBSERVADOS Y AJUSTADO PARA COMBUSTIBLE, KERNEL POLINOMIAL DE ORDEN 2 (TEST) PARA SVM</i>	101

<i>FIGURA 75 - COMPARACIÓN VALORES OBSERVADOS Y AJUSTADO PARA EL CONSUMO DE COMBUSTIBLE CON SVM.....</i>	<i>102</i>
<i>FIGURA 76 - DIAGRAMA DE CAJAS DEL ERROR RELATIVO CON 218 OUTLIERS PARA AJUSTE DE COMBUSTIBLE-SVM</i>	<i>102</i>
<i>FIGURA 77 – RMSE (ENTRENAMIENTO) EN FUNCIÓN DE CANTIDAD DE NEURONAS PARA ESTIMADO DE CONSUMO DE GASOLINA.</i>	<i>105</i>
<i>FIGURA 78 - EVOLUCIÓN DE ERROR RELATIVO ESTIMADO DE CONSUMO DE COMBUSTIBLE PARA DATOS DE TEST DE LAS TABLAS 23, 24 Y 25.....</i>	<i>107</i>
<i>FIGURA 79 - RMSE PARA 1 NEURONA A 20 NEURONAS ESTIMADO DEL CONSUMO DE COMBUSTIBLE. .</i>	<i>107</i>
<i>FIGURA 80 - MSE PARA ENTRENAMIENTO, VALIDACIÓN Y TEST CON 19 NEURONAS PARA ESTIMADO DE CONSUMO DE GASOLINA</i>	<i>108</i>
<i>FIGURA 81 - ENTRENAMIENTO DE LA RED CON 17 NEURONAS PARA CONSUMO DE COMBUSTIBLE</i>	<i>109</i>
<i>FIGURA 82 - VALORES OBSERVADOS VERSUS PREDICCIÓN PARA CONSUMO DE COMBUSTIBLE CON 19 NEURONAS EN LA CAPA OCULTA.....</i>	<i>109</i>
<i>FIGURA 83 - DIAGRAMAS DE CORRELACIÓN PARA PREDICCIÓN DE CONSUMO DE COMBUSTIBLE, PARA ENTRENAMIENTO (IZQUIERDA) Y TEST (DERECHA).....</i>	<i>110</i>
<i>FIGURA 84 - DIAGRAMA DE CAJAS DE ERROR RELATIVO (PARA TEST) CON 113 OUTLIERS PARA ESTIMADO DEL CONSUMO DE COMBUSTIBLE-RN CON 19 NEURONAS EN LA CAPA OCULTA.</i>	<i>110</i>

NOMENCLATURA

<i>AAP</i>	Asociación Automotriz del Perú
<i>CONAM</i>	Concejo Nacional de Medio Ambiente
<i>CARB</i>	<i>California Air Resources Board</i>
<i>EPA</i>	<i>Enviromental Protection Agency</i>
<i>ETSII</i>	Escuela Técnica de Ingeniería Industrial. Universidad Politécnica de Valencia
<i>ETM</i>	Empresa de transporte municipal
<i>INSIA</i>	Instituto nacional de investigación del automóvil
<i>LH</i>	Litro por hora
<i>MAE</i>	Error Absoluto Medio
<i>MAPE</i>	Error Porcentual Absoluto Medio
<i>MPG</i>	Millas por galón
<i>MSE</i>	Error cuadrático medio (Mean Squared Error)
<i>NMSE</i>	Error cuadrático medio normalizado (<i>Normalised Mean Squared Error</i>)
<i>NMAE</i>	Error Absoluto Medio Normalizado
<i>OBDII</i>	<i>On Board Diagnostics II</i>
<i>RMSE</i>	Raíz cuadrada del error cuadrático medio
R^2	Coefficiente de determinación
<i>RPM</i>	Revoluciones por minuto
<i>RNN</i>	<i>Recurrent Neural Networks</i>
<i>SVM</i>	<i>Support Vector Machine</i> (Máquina de soporte vectorial)
<i>SST</i>	Suma de cuadrados totales

SSR	Regresión de la suma de cuadrados
SSE	Error de suma de cuadrados
VE	Varianza explicada
UE	Unión Europea
y_i	Valores observados o reales
\hat{y}_i	valores estimados
\bar{y}	El valor medio de los y_i

VARIABLES USADAS

T	Variable objetivo
trn_T	Variable objetivo de entrenamiento
$test_T$	Variable objetivo de prueba
XY	Variable matriz
X	Variable predictora
trn_X	Variable predictora de entrenamiento
$test_X$	Variable predictora de prueba
$test_T_pred$	Predicción variable objetivo de prueba
trn_T_pred	Predicción variable objetivo de entrenamiento
$idxTrn$	Índice de conjunto de entrenamiento
$idxTest$	Índice de conjunto de entrenamiento
n	número de observaciones
p	número términos de la regresión incluido el término interceptor

INTRODUCCIÓN

En el presente estudio se estima el consumo de gasolina y las emisiones de CO₂ de un motor vehicular, mediante técnicas de aprendizaje automático (Machine Learning) como alternativa a los métodos estadísticos actuales. Las técnicas utilizadas son tres: Regresión Múltiple, Máquina de Soporte Vectorial y Redes Neuronales. Una de las características de estas técnicas es que permiten analizar grandes volúmenes de datos y su procesamiento significa un ahorro de tiempo frente a otros.

La estimación de estas variables de estudio es importante debido a la creciente preocupación por el aumento de la demanda de combustibles fósiles en el transporte y el calentamiento global, conllevando a la necesidad de implementar límites, a partir de estimaciones fiables.

El uso de combustibles fósiles está en constante aumento, por varias causas, entre las que están el crecimiento poblacional y, consecuentemente, el crecimiento del parque automotor en las naciones. Tal es el caso de China, quien ha anunciado que alcanzará el punto máximo de emisiones de CO₂ el 2030 y la neutralidad en 2060 [1] (DOI: 10.1016/j.scib.2021.10.022). Así mismo, un reciente estudio indica que al 2030, la economía del país asiático tendría un PBI real de 6.8% menor al que tendría sin cambio climático [2].

Según un informe del diario El Peruano [3], se indica que la venta de vehículos eléctricos se limitará al 5% en el 2030, debido a diferentes factores entre los cuales está el elevado costo comparado a los vehículos convencionales. Desde enero 2022 a agosto 2022, se vendieron 105 163

vehículos convencionales y el año anterior 96 468 unidades. Para el caso de vehículos eléctricos, para el mismo periodo, se vendieron un total de 2 268 unidades [4]. Por lo tanto, bajo este escenario actual, el incremento de gases de efecto invernadero del transporte seguirá en aumento. Además, de acuerdo con InfraLatam y el BID (2019) el sector transporte nacional representó el 5.32 % del PBI en el 2015, uno de los más altos en América del Sur [2]

En el parque automotor nacional existen diferentes categorías y tecnologías vehiculares, pero el estudio se centra en un motor de vehículo ligero, ya que estos representan un porcentaje importante del parque automotor, según los datos de importación del MTC. De los 175 227 vehículos importados el 2021, los vehículos ligeros (automóviles, Station Wagon y camionetas) representaron el 89.5% del total [5].

Por lo anteriormente expuesto, se justifica la creación de una propuesta metodológica para la construcción de modelos, basado en procedimientos de obtención de datos, tratamiento y creación, mediante Machine Learning. La mejor predicción del consumo de combustible y emisiones de CO₂, utilizando datos reales obtenidos del motor a gasolina instalado en un banco de pruebas, permitirán mejorar la estimación del comportamiento de estas variables de estudio, bajo condiciones de funcionamiento del motor en el transporte urbano de Lima. Además, el estudio de tesis es una propuesta metodológica para utilizar e identificar las mejores técnicas de Machine Learning para uso en categorías de motores y vehículos de transporte.

Definición del problema

La estimación del consumo de combustible y emisiones de CO₂ generadas por un motor a gasolina, a partir de datos de su funcionamiento, requiere ser contrastada. La estimación de estas variables, asociadas, es fundamental por varias razones, entre las que destaca la sostenibilidad y el medio ambiente, ya que buscan reducir el consumo de gasolina y las emisiones de gases de efecto invernadero.

Las técnicas de Machine Learning permiten estimar estas variables. Sin embargo, requiere de una gran cantidad de datos y la implementación de varias técnicas para efectos de comparar sus predicciones.

Propuesta de Solución

Desarrollar pruebas experimentales sobre un motor a gasolina y recolectar datos de funcionamiento para implementar tres técnicas de Machine Learning para estimar el consumo de combustible y emisiones de CO₂.

La fuente de datos de motor proviene del módulo de control electrónico del motor y sensores, además de los equipos de medición instalados en el banco de pruebas. Los datos serán procesados mediante Regresión Múltiple, Máquina de Soporte Vectorial y Redes Neuronales. Las estimaciones de las tres técnicas de Machine Learning serán analizadas. Además, se evaluará la minimización de errores al realizar interacciones entre ellas.

Justificación

La importancia del uso de Machine Learning en modelos predictivos sobre emisiones de CO₂ y consumo de combustible, a partir de datos del motor, se justifica porque existen varias variables predictoras y cuya ecuación de regresión tiene términos no lineales. Las técnicas de Machine Learning permiten realizar modelos que se ajustan a diversos problemas, sin la necesidad de ajuste de variables, sí lo requieren ciertos modelos estadísticos. Estos modelos se entrenan para ajustar y predecir el resultado esperado; para ello, toma parte de los datos y reserva otra para validación y prueba. Los datos y modelos pueden ser luego de consulta y pueden ser utilizados en conjunto con una base de datos más amplia de motores de vehículos de otras categorías, para un análisis más general.

Otra aplicación de importancia de un modelo de regresión es que, ajustado con datos de un motor en buen estado de funcionamiento, se puede predecir el consumo de combustible y emisiones de CO₂, que pueden ser contrastados con mediciones de sensores del motor y detectar un consumo elevado y emisiones fuera del rango permitido; y de esta manera poner sobre aviso algún desperfecto.

La aplicación de las Técnicas de Machine Learning es viable debido a que los datos necesarios son accesibles mediante pruebas del motor en un banco dinamométrico (cumplen con el requisito de repetibilidad), y los modelos de Machine Learning se pueden aplicar utilizando diversos programas de alto nivel disponibles, algunos incluso con licencia de software libre.

Objetivos

Objetivo general

Estimar el consumo de combustible y las emisiones de CO₂ en un motor de 130 HP a gasolina mediante técnicas de Machine Learning: Regresión Múltiple, Máquina de Soporte Vectorial y Redes Neuronales

Objetivo específico

- Elaborar un procedimiento para obtener datos de funcionamiento de un motor a gasolina instalado en banco de pruebas.
- Realizar pruebas del motor a gasolina en distintas condiciones de torque y régimen de giro. Recolectar datos de sensores e instrumentos, como del medidor de consumo de combustible y analizador del CO₂ en los gases de escape.
- Procesar los datos mediante las técnicas de Regresión Múltiple, Máquina de Soporte Vectorial y Redes Neuronales.
- Comparar las estimaciones de las diferentes técnicas de Machine Learning y analizar la contribución de las diferentes interacciones entre ellas, buscando la minimización de errores.

CAPÍTULO 1

FUNDAMENTO TEÓRICO

1.1. Scanner OBD

OBD (On Board Diagnostics) es un sistema de diagnóstico a bordo en vehículos (coches y camiones), que apareció con la finalidad de reducir las emisiones contaminantes de los vehículos [6].

A principios de la década de 1970, los vehículos vendidos en los Estados Unidos se equiparon con la electrónica para el control de varios sistemas y diagnosticar fallos del coche, con el objetivo de reducir al mínimo la contaminación. Esto se produjo al aprobarse la Ley de Aire Limpio y el establecimiento de la Agencia de Protección Medioambiental (EPA), en 1970. Esta electrónica variaba entre los fabricantes y el año del modelo [7].

En el año 1991 “California Air Resources Board” (CARB) pone como requisito que todos los automóviles vendidos en California tengan un sistema de OBD, este sistema implementó unas luces indicadoras para que el conductor se informe de un fallo en el automóvil. No era un estándar ni en el equipo ni el protocolo de comunicación, y luego se conoció como OBD-I. Años más tarde, CARB, en California, obligó a vehículos a contar con un OBD-II junto con los protocolos de SAE. Tiempo después la Unión Europea impuso de forma obligatoria que todo vehículo gasolinero desde el año 2000 y vehículo Diesel, desde el año 2003, debían contar con este protocolo.

El sistema OBD-II puede detectar fallos en los diferentes sensores con que los automóviles cuentan hoy en día. Las formas en que el sistema informa sobre los fallos posibles son a través de luces indicadoras (*Malfunction Indication Lamp* (MIL), ó *Check Engine*) y también a mediante códigos que son registrados en el momento en el que se producen y que pueden ser leído por medio de un scanner. Uno de ellos es el scanner que se muestra en la Figura 1, el ELM 327 que se comunica con un Smartphone por medio de *bluetooth*.



Figura 1 - Interface ELM327 [3].



Figura 2 - Descripción de terminales OBD2.

1.2. Machine Learning

1.2.1. Regresión Líneal Multiple

Un modelo de regresión describe la relación entre variables dependientes, y , y una o más variables independientes, X . Las variables dependientes también son llamadas variables respuesta u objetivo y las variables independientes también son llamadas variables predictoras.

$$y_j = \beta_0 + \beta_1 x_{i1} + \beta_2 x_{i2} + \dots + \beta_p x_{ip} + \varepsilon_j, \quad i = 1, \dots, n, \quad (1)$$

donde:

- y_j i-ésima respuesta;
- β_k es el k-ésimo coeficiente;
- X_{ij} es la i-ésima observación sobre la j-ésima variable predictora, $j = 1, \dots, p$;
- ε_i es el i-ésimo termino ruido que es el error aleatorio.

Si el modelo incluye solo una variable predictora ($p = 1$) entonces el modelo es llamado modelo de regresión lineal simple:

$$\hat{y}_j = \sum_{k=0}^K B_k f_k(X_{i1}, X_{i2}, \dots, X_{ip}), \quad i = 1, \dots, n, \quad (2)$$

\hat{y}_j es llamado respuesta estimada y B_k s son los coeficientes de ajuste, estos son estimados para minimizar la diferencia de los cuadrados medios entre el vector ajustado y el verdadero valor del vector respuesta $\hat{y}_j - y$. Este método es llamado método de mínimos cuadrados. Bajo los supuestos de los

términos de ruido; estos coeficientes también maximizan la probabilidad del vector de predicción [8].

Previo a realizar el ajuste de regresión se tiene que preparar los datos y esta fase tiene varios pasos, que inician con eliminar valores nulos, es decir la recolección de datos puede por diferentes razones registrar valores nulos en varios campos de la variable predictora como de la variable respuesta. El siguiente paso es elegir el modelo que para el caso están clasificados en Matlab como:

- 'constant': modelo que contiene solo un término;
- 'linear': contiene un término lineal y un interceptor para cada predictor;
- 'interactions': contiene un término lineal y un interceptor y todos los productos de pares de distintos predictores (sin incluir términos cuadráticos);
- 'purequadratic': moldeo que contiene un interceptor, términos lineales y términos al cuadrado;
- 'quadratic': contiene un interceptor, términos lineales, interacciones y términos cuadráticos;
- 'polyijk': contiene un polinomio con todos los términos hasta el grado i en el primer predictor, grado j en el segundo, etc., se puede usar número del 0 al 9 por ejemplo, 'poly2111' contiene términos constantes más todos los términos lineales y productos y también tiene términos con el predictor 1 elevado al cuadrado.

Una herramienta útil para realizar un ajuste de regresión es la distancia de Cook, ésta es el cambio escalado en valores ajustados, el cual es útil para

identificar valores atípicos en el valor de la variable X (observación de la variable predictora X). La distancia de Cook muestra la influencia de cada observación en el valor. Una observación con un “Cook’s distance” mayor a tres veces la media de Cook’s distance puede ser un valor atípico, conforme se aprecia en la Figura 3 [9].

Podemos hacer el ajuste luego de identificar los valores típicos y retirarlos para nuevamente realizar la regresión y mejorar nuestro resultado si es necesario:

$$D_i = \frac{\sum_{j=1}^n (\hat{y}_j - \hat{y}_{j(i)})^2}{pMSE} \quad (3)$$

donde:

\hat{y}_j es el j-ésimo valor de respuesta ajustado;

$\hat{y}_{j(i)}$ es el j-ésimo valor de respuesta ajustado, donde el ajuste no incluye el valor i;

MSE El error cuadrático medio;

p Es el número de coeficientes en el modelo de regresión.

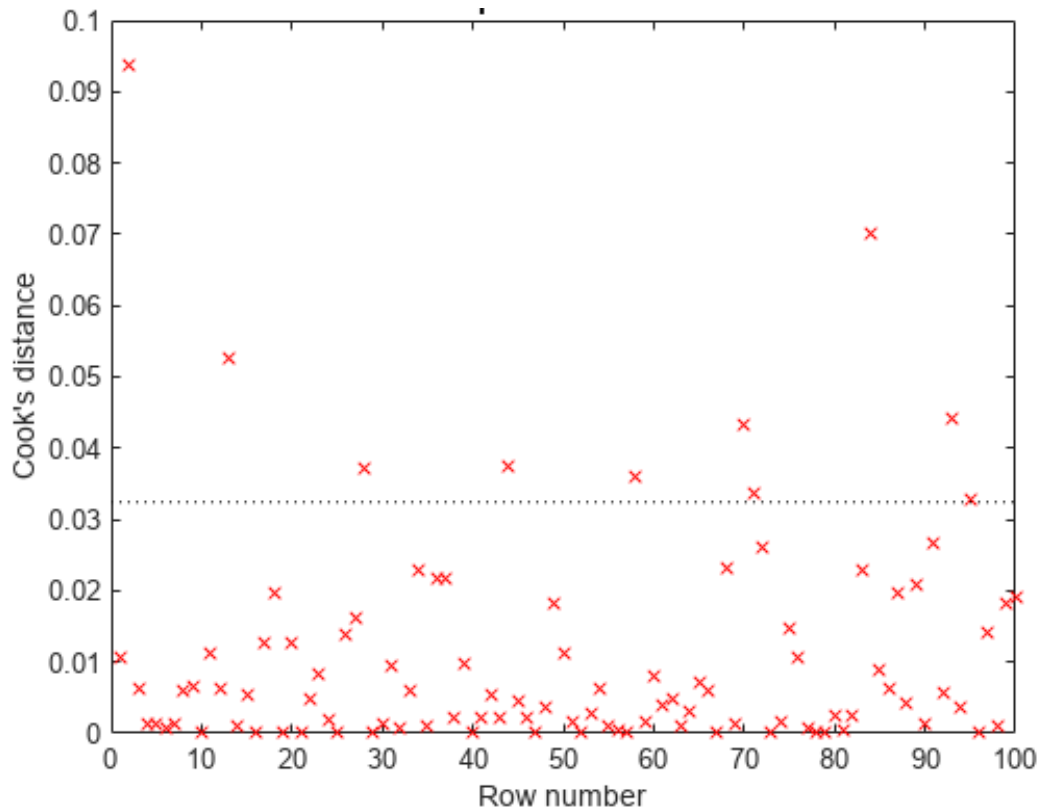


Figura 3 - Cook'sDistance-Tomado de MatWorks [5]

Otra herramienta útil es el apalancamiento (*Leverage*), que es la medida del efecto de una observación sobre la predicción de una regresión debido a la posición de la observación en el espacio de las entradas. En general, cuanto más lejos está un punto del centro del espacio de entrada, más apalancamiento (*Leverage*) tiene. Como la suma de los valores de apalancamiento es p , una observación i puede ser considerada como un valor atípico si su *Leverage* excede sustancialmente su valor medio, p/n ; por ejemplo un valor mayor que $2 \cdot p/n$ (véase Figura 4) [10].

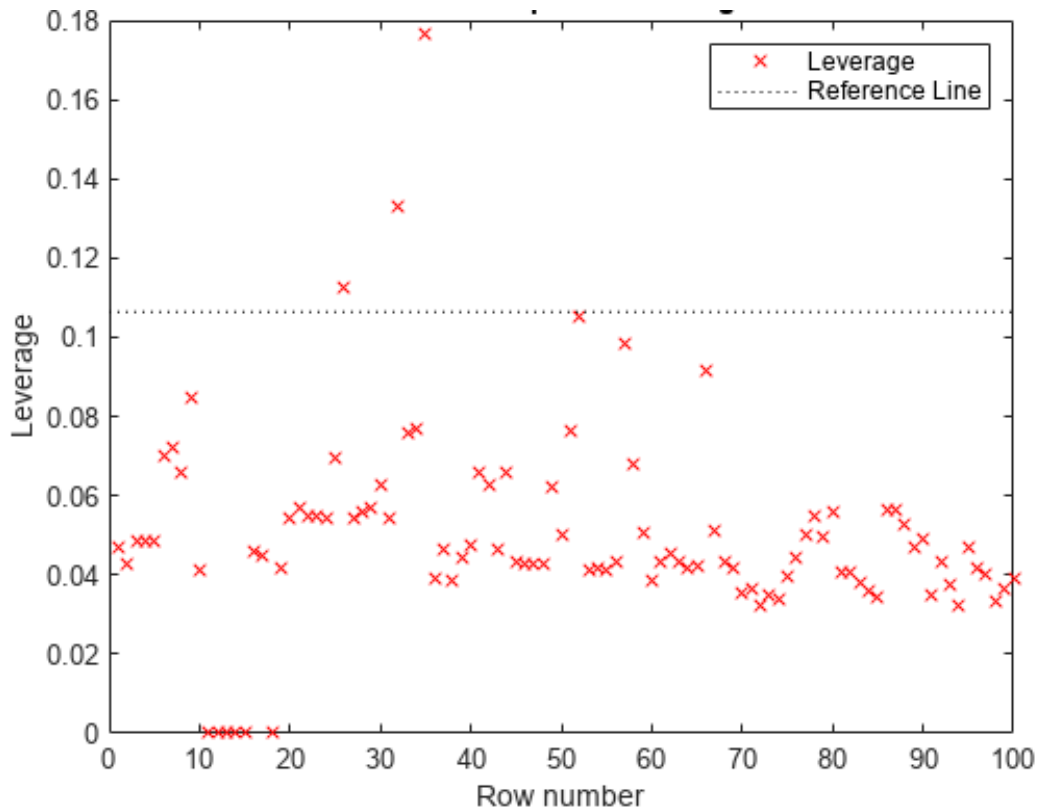


Figura 4 - Leverage-Tomado de MatWorks [10]

1.2.2. Máquina de Soporte Vectorial

Es un algoritmo que está clasificado como supervisado, que se basa en la construcción de un hiperplano que separe los datos para poder ser clasificados de manera eficiente y, de esta manera, se genere un modelo que puede ser usado con datos nuevos y poder clasificarlos. También puede ser usado para regresión. En casos que no sea posible la separación mediante un hiperplano se puede realizar una transformación usando funciones llamadas kernel, es decir. Por ejemplo, tenemos dos tipos de datos en el plano que no pueden ser separados por una recta, podemos, en este caso, usar una función kernel para transformar los datos de dos dimensiones y llevarlos a

una dimensión superior en la que podemos separarlo mediante un hiperplano.

Existen varios tipos de *kernels* y el más usado es el Gaussiano.

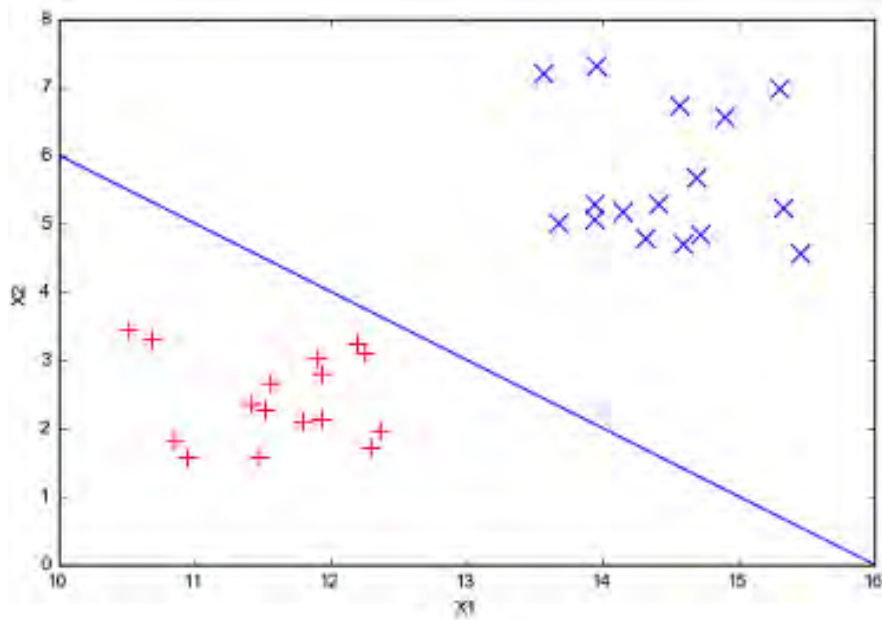


Figura 5 - Datos separados por una línea [11].

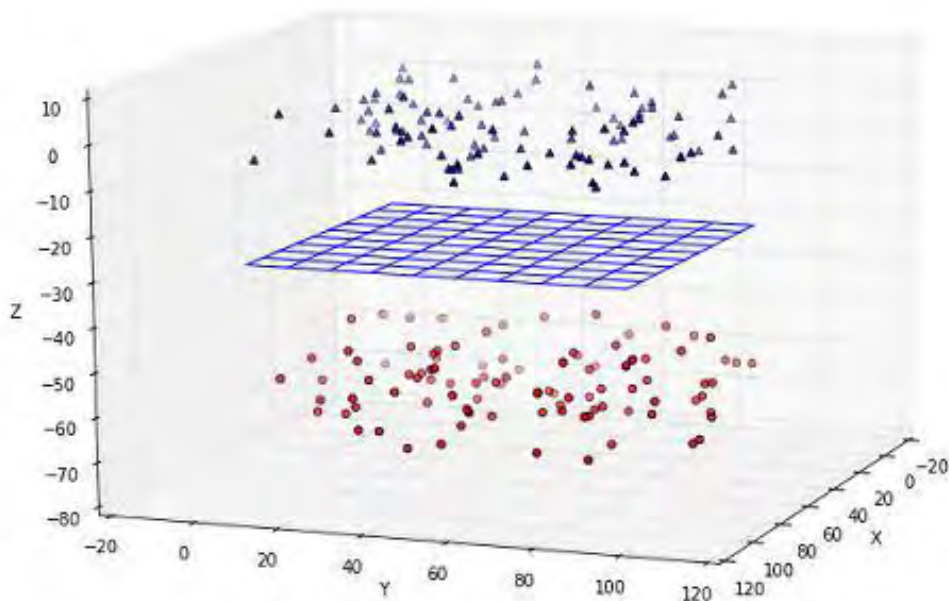


Figura 6 - Datos separados por un plano [11].

1.2.3. Redes neuronales

Inteligencia artificial simbólica y subsimbólica. Simbólica implica que debemos conocer una aproximación a la solución del problema y diseñarla completamente con el aporte de conocimiento de un experto sobre el tema. Por otra parte, se encuentra la subsimbólica en donde no se realiza diseños de alto nivel, sino que, por si solo el sistema va adaptándose y generando un sistema capaz de resolver el sistema.

La perspectiva subsimbólica trata de estudiar los mecanismos de los sistemas nerviosos, del cerebro, así como su estructura, funcionamiento y características lógicas, con la intención de diseñar programas basados en dichas características que se adapten y generen sistemas capaces de resolver problemas. En este caso el sistema es de abajo hacia arriba (botón – up). Ya que los sistemas son simples e idénticos, recogen las características físicas de los sistemas que tratan de imitar y se van generando cálculos cada vez más complejos, de forma automática, mediante mecanismos prefijados. Es en este campo donde se encuadran las Redes de Neuronas Artificiales [12].

La estructura de las Redes Neuronales está formada por múltiples neuronas organizadas en múltiples capas, las que están agrupadas al inicio, reciben el nombre de capas de entrada, luego vienen las intermedias o capas ocultas y las de salida. Las conexiones entre capas tienen asociadas pesos W y funciones de activación F , es decir al recibir una señal en la entrada que puede ser un vector con múltiples valores, estos son multiplicados por pesos cuyo resultado son señales E ; estas señales, a su vez, son procesadas por una función de activación F que produce la señal de salida S . Hay múltiples

funciones de salida, el proceso se repite con cada capa adicional dependiendo del modelo o estructura de la red.

Dentro de las funciones de activación existen funciones lineales y no lineales, estas últimas son las que dan a la red el potencial para resolver problemas de forma genérica con más éxito.

El tipo de datos tomados para entrenar una red debe cumplir con ser representativa es decir los datos deben ser diversos. Todos los datos deben estar representados en el conjunto de aprendizaje. Y también deben ser significativos, es decir debe haber un número suficiente de ejemplos para que la red pueda adaptarse para resolver los problemas.

Para el caso de redes usadas en regresiones, se tiene un número de ciclos decidido antes y también está el valor del error establecido. Una vez concluido el número de veces, indicado por la cantidad de ciclos, se da por aceptado la red resultante y puede darse el caso que el programa pare antes de cumplir en número de ciclos cuando el error descienda por debajo del valor preestablecido. Es decir, el final puede alcanzarse por estas dos condiciones: primero por el valor del error; y luego por el número de ciclos alcanzados.

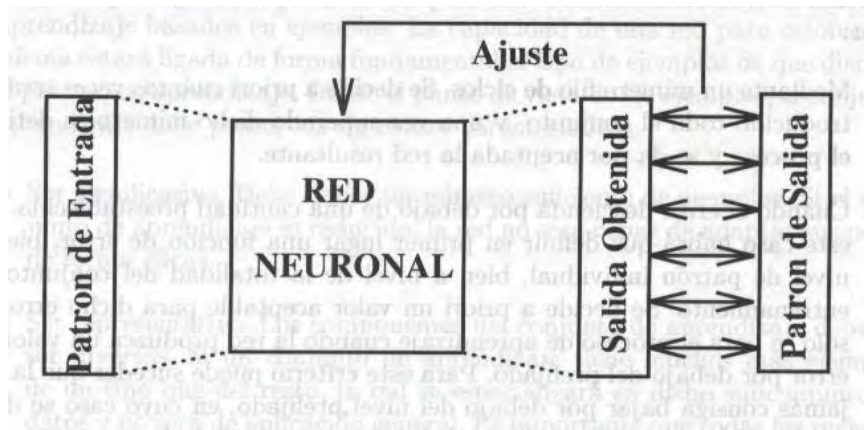


Figura 7 - Aprendizaje supervisado. Fuente: redes neuronales artificiales [12].

Otro criterio para terminar con el proceso, es el check de validación, el cual verifica que cumplidos el máximo de iteraciones consecutivas (por defecto 6) en los cuales el error de validación medido por MSE no disminuye.

Las iteraciones de datos completos se miden en épocas. Esta cuenta las veces que todos los datos son usados para el entrenamiento hasta que se alcance los criterios de validación para los cual el valor por defecto si no se especifica es de 1000 épocas para Matlab, en el caso de redes pocos profundas que usan el algoritmo de entrenamiento “Levenberg-Marquardt”, “Retro propagación de regularización Bayesiana” y “Retro propagación de gradiente conjugado escalado”.

A continuación, se presenta una breve descripción de estos algoritmos de entrenamiento y sus valores por defecto:

- **Levenberg-Marquardt:** es un algoritmo rápido, que utiliza un arreglo con matrices Jacobianas para aproximar la matriz hessiana. $x_{k+1} =$

$x_k - [J^T J + \mu I]^{-1} J^T e$, donde J es la matriz jacobiana. Si $\mu = 0$, transforma a la expresión anterior en el método de Newton [13].

- **Retro propagación de regularización Bayesiana:** usa como base el algoritmo Levenberg -Marquardt. La retro propagación se usa para calcular la matriz Jacobiana del rendimiento respecto a las variables de peso y sesgo. Cada variable se ajusta según Levenberg-Marquardt [14].
- **Retro propagación de gradiente conjugado escalado:** con valores de 1000 épocas, objetivo de rendimiento 0, máximo errores de validación 6, μ predeterminado 0.005. Usa peso, entrada de red y funciones de transferencia tengan funciones derivadas

A seguir, se muestran las funciones de activación más comunes usadas en el diseño de redes neuronales [15]:

$$\text{hardlims} = 1 \text{ if } n \geq 0, -1 \text{ de otra forma.} \quad (4)$$

$$\text{hardlim} = 1 \text{ if } n \geq 0, 0 \text{ de otra forma.} \quad (5)$$

$$\text{logsig} = f(x) = \frac{1}{1+e^{-x}} \quad (6)$$

$$\text{purelin} = a = \text{purelin}(n) = n \quad (7)$$

$$\text{tansig} = f(x) = \frac{2}{1+e^{-2x}} - 1 \quad (8)$$

En el presente trabajo la función “Tansig” en la capa oculta y “Purelin” en la capa de salida de la red; “Hardlim” es usada en la red “Perceptron” y “Purelin” es usada en “Adaline”.

Una de las primeras redes fue el “Perceptron” y tiene la siguiente arquitectura [15] mostrada en la Figura 8. Es una red de una sola capa con R entradas y S neuronas, con pesos W de S x R y bias (sesgo) de Sx1, con una función de activación “Hardlim”. En esta red cada dato de entrada tiene asociada una salida (ver Figura 8).

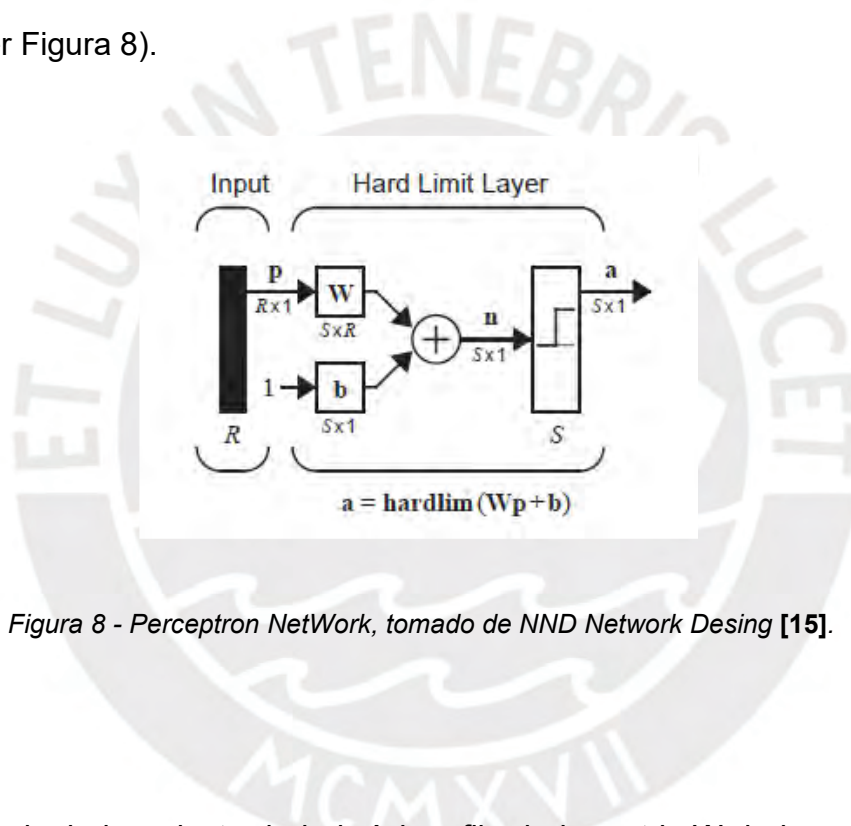


Figura 8 - Perceptron NetWork, tomado de NND Network Desing [15].

En esta red, si el producto de la i-ésima fila de la matriz W de los pesos con el vector de entrada P es mayor o igual a -b, la salida será 1; de otra forma, será 0. Una red neuronal “Perceptron” de varias neuronas puede realizar una clasificación en tantas salidas como 2^s , donde s es el número de neuronas de la red. Sin embargo, esta red de una sola capa solo puede resolver problemas que son linealmente separables.

La red “Feedforwardnet”, es una red neuronal alimentada con valores predeterminados. Por ejemplo, el comando para crear una red con 10 neuronas, sería: `net=feeforwardnet (10)`.

La Figura 9 muestra una Red con dos capas, una capa oculta con 10 neuronas y otra capa de salida con una, la primera y segunda capa con pesos y bias. Una función de activación ‘Tansig’ en la capa oculta y una función lineal ‘Purelin’ en la capa de salida.

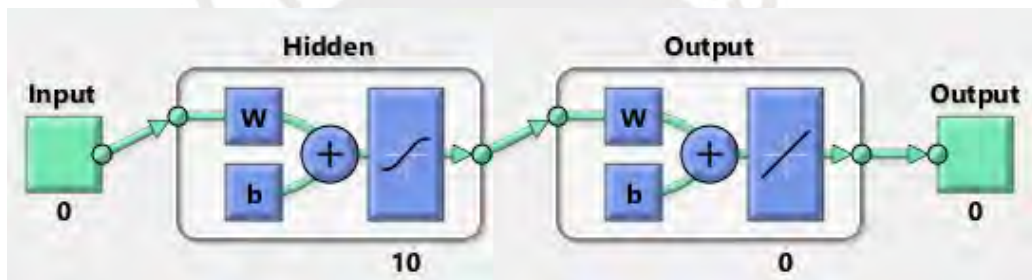


Figura 9 - Red Neuronal, elaboración propia.

1.3. Fórmulas para el análisis

- Suma de cuadrado debido al error:

$$SSE = \sum_{i=1}^N (\hat{y}_i - y_i)^2 \quad (9)$$

- Suma de cuadrados de regresión:

$$SSR = \sum_{i=1}^N (\hat{y}_i - \bar{y})^2 \quad (10)$$

- Suma total de los cuadrados:

$$SST = \sum_{i=1}^N (y_i - \bar{y})^2 \quad (11)$$

- R-Cuadrado ordinario:

$$R^2 = \frac{SSR}{SST} = 1 - \frac{SSE}{SST} \quad (12)$$

- R-Cuadrado ajustado al número de coeficientes:

$$R^2 = 1 - \left(\frac{n-1}{n-p} \right) \frac{SSE}{SST} \quad (13)$$

Las Fórmulas (10), (11), (12) y (13) fueron tomadas de la referencia [16].

- Coeficiente de determinación múltiple:

$$R^2 = 1 - \frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N (y_i - \bar{y})^2} \quad (14)$$

- Error absoluto medio:

$$MAE = \frac{1}{N} * \sum_{i=1}^N |y_i - \hat{y}_i| \quad (15)$$

- Error absoluto medio normalizado:

$$NMAE = \frac{\sum_{i=1}^N |y_i - \hat{y}_i|}{\sum_{i=1}^N |y_i|} \quad (16)$$

- Raíz del error cuadrático medio

$$RMSE = \sqrt{\frac{1}{N} * \sum_{i=1}^N (y_i - \hat{y}_i)^2} \quad (17)$$

- Error cuadrático medio

$$MSE = \frac{1}{N} * \sum_{i=1}^N (y_i - \hat{y}_i)^2 \quad (18)$$

- Error cuadrático medio normalizado

$$NRMSE = \sqrt{\frac{\sum_{i=1}^N (y_i - \hat{y}_i)^2}{\sum_{i=1}^N y_i^2}} \quad (19)$$

donde y_i son los valores observados o reales; \hat{y}_i valores estimados; \bar{y} el valor medio de y_i .

En el presente trabajo de tesis se utilizarán, principalmente, los estadísticos RMSE, MSE, R^2 .

CAPÍTULO 2

ESTADO DEL ARTE

Un modelo de regresión para emisiones de CO₂ aplicada a LDDTs (Light-Duty Diesel Trucks) publicada en 2022 por Zhang y colegas, en la revista “Atmosphere” [17] (DOI: 10.3390/atmos13091466), muestra el uso de “Deep Learning” (técnica de Machine Learning) en la elaboración de un modelo para predecir emisiones de CO₂ basado en LDDT llamado DL-DTCEM (Deep Learning-Based LDDT CO₂ Emission Model) que usa como variables las captadas mediante PEMS (Portable Emission Measurement System) y GPS (Global Position System). En ese artículo se hace una diferencia en el análisis de acuerdo al alcance en categorías de modelos: el modelo macroscópico, que usa parámetros como velocidad promedio; y un modelo microscópico que toma parámetros instantáneos de conducción del vehículo como velocidad y aceleración para estimar la tasa de emisiones de contaminantes. Los autores mencionan que los métodos estadísticos han sido por años los más usados en la elaboración de modelos, uno de los cuales es el CMEM (Comprehensive Modal Emissions Model), que usa modelos polinomiales con diferentes parámetros como son coeficiente aerodinámico, masa del vehículo, control tecnológico de emisiones, sistema de distribución de combustible. Pero, según apreciación de los investigadores, la ausencia de información en algunos casos puede llevar a falta de precisión en la predicción. Este estudio tomó tres escenarios: autopistas, zona urbana y zona suburbana. Para datos de aceleración tomo el promedio de instante previo y siguiente. Otro dato fue el grado o diferencia de altitud como la sumatoria de altitud cada 50 m. Un

dato adicional es el VSP que está en función de la velocidad, aceleración y el grado. La división de datos para entrenamiento y prueba del modelo fue 80% y 20%, respectivamente. Utilizaron un modelo de Deep Learning con una capa densa con funciones de activación Sigmoid y Tanh. La evaluación la realiza utilizando RMSE, haciendo una comparación entre valor real (RDE, Real Driving Emission) y predicción. También analiza el incremento de emisiones de CO₂ para VSP con valores mayores que cero. Finalmente, muestra resultado de dos casos, uno con un vehículo con estándar de emisiones CHINA III y otro con CHINA IV, ambos realizados con un mismo conductor para evitar diferencias en el patrón de conducción. El modelo alcanza un valor de R² y un RMSE para ambos casos de 0.986-0.990 y 0.165-0.167, correspondientemente.

La Figura 10 muestra los datos para valores observados y valores estimados por el modelo, con datos registrados en diferentes vehículos, pero con un mismo conductor. Los autores concluyen que las emisiones de CO₂ aumentan rápidamente cuando la velocidad se incrementa más de 5m/s y la aceleración 0.5m/s². Aun cuando este estudio se realiza en dos tipos de vehículos típicos, estos modelos pueden ser ajustados a otros tipos de vehículos en futuros estudios

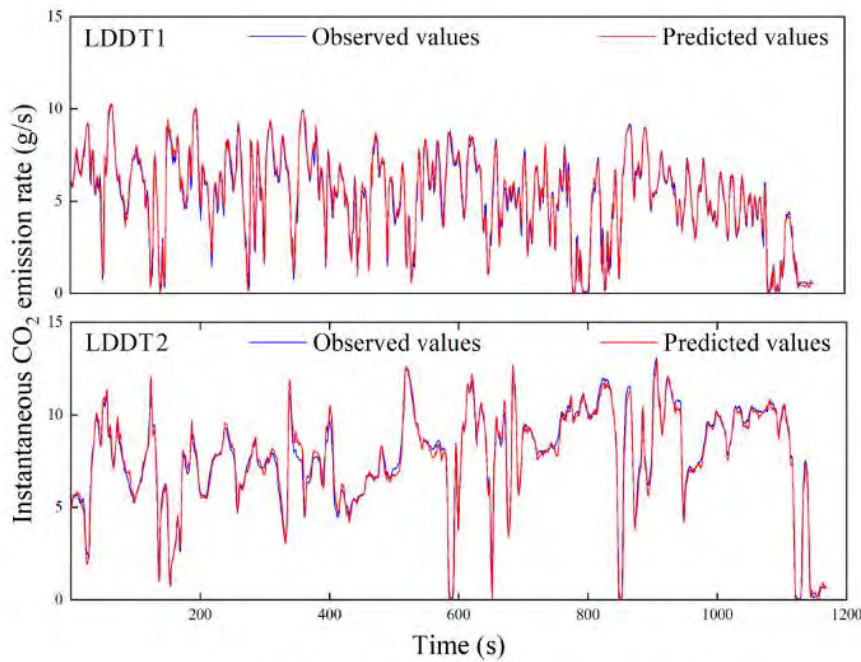


Figure 6. Comparison of DL-DTCM predicted CO₂ emission rate with observed data.

Figura 10 - Emisiones de CO₂ reales y predicho. Tomado de referencia [17].(DOI: 10.3390/atmos13091466).

El modelo de emisiones realizado por Madziel *et al.* [18], (DOI: 10.3390/en15010142), publicado en 2022 en la revista “Energies” muestra el uso de técnicas de Machine Learning para desarrollar un modelo para predecir emisiones instantáneas como alternativa a la inviable instalación de equipos en vehículos para medición de emisiones. También menciona que el uso de modelos actuales no es viable debido a que son usados en macro escala, por lo tanto, es necesario el desarrollo de modelos para microescala, el tipo de vehículos analizados con híbridos.

El incremento del parque automotor de vehículos híbridos hace necesario contar con modelos para predecir emisiones, y en la literatura, según los autores, no hay suficientes modelos para esta categoría de vehículos híbridos

para lo que usó diferentes técnicas de Machine Learning como Regresión Lineal, Tree, SVM, GPR (Gaussian process regression) y NNET (neural network). El método que dio mejores resultados fue el GPR. La recolección de información se realizó mediante PEMS system, y el vehículo fue evaluado con un analizador de gases para asegurar que el valor de las emisiones esté dentro de los rangos normales. Las características del vehículo fueron 1497cm³, año de fabricación 2020, y cumple estándar de emisiones EURO 6. La prueba se realizó en tres escenarios: autopista, rural y urbano. El modelo fue realizado utilizando el “Matlab’s Regression Learner application” que permite utilizar el mejor método disponible, se utilizó 5 “fold cross validation”. Los datos son partidos aleatoriamente en 5 partes, 4 conjuntos de datos para construcción del modelo y una el conjunto de prueba. Para analizar el resultado utilizó el coeficiente de determinación RMSE y MSE. El mejor $R^2 = 0.69$ obtenido fue para el modelo GPR y el peor para el SVM, con un valor de $R^2 = 0.41$.

La Figura 11 muestra la comparación entre valores reales y valores realizados por el modelo de predicción realizado. Los autores concluyen que las emisiones de CO₂ en vehículos híbridos no están estrictamente relacionados con la velocidad porque en algunos rangos de operación estos vehículos no emiten.

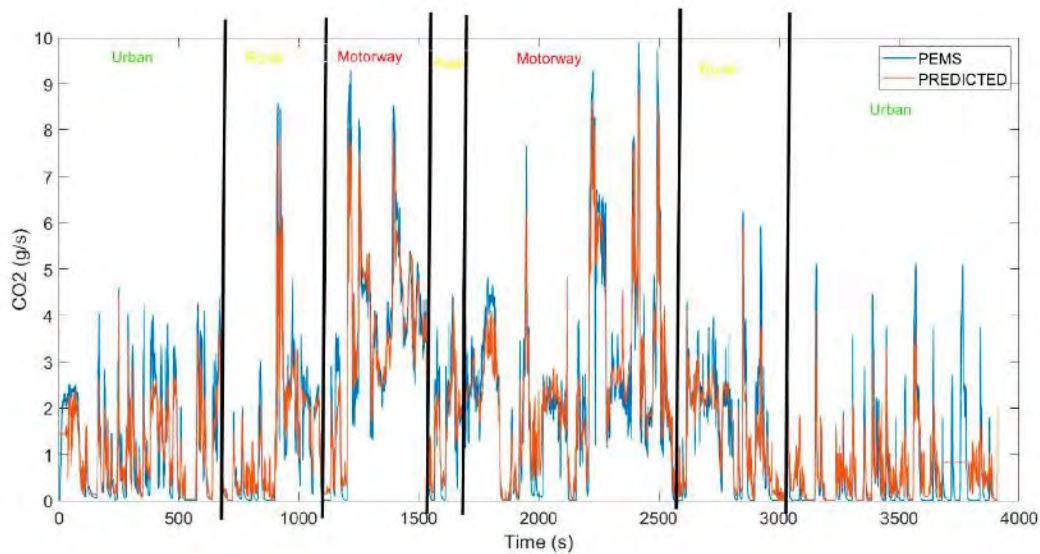


Figura 11 - Emisiones de CO₂, real y predicción. Tomado del modelo de emisiones publicado por Madziel y colegas [18].(DOI: 10.3390/en15010142).

Caro, en 2021, realizó un estudio [19] para estimar el consumo de combustible en vehículos. En su estudio utilizó dos ciclos de manejo para toma de datos: el de manejo NEDC, utilizado por Volkswagen y cuyo caso “dieselgate” fue conocido mundialmente; y el ciclo WLTP, que tiene una mejora, pero permite que los fabricantes puedan usarlos para salir bien en dichas pruebas. Por ello, además planteó una metodología más fiable para estimar el consumo de combustible, proponiendo dos posibilidades; un modelo tipo caja blanca que tomaba datos del motor tomados mediante el puerto OBDII; y otro modelo tipo caja negra donde solo tomó datos de velocidad y altitud tomados por un GPS. Al final decidió tomar solo datos del GPS y con esto realizó su estudio en tres escenarios:

- Básico: velocidad instantánea y variación respecto a t+1 y altitud respecto a t+1;

- Medio: velocidad y variación de velocidad respecto a, $t-5$, $t-3$, $t-2$, $t-1$, $t+1$, $t+2$, $t+3$ y $t+5$, y las variaciones de altitud respecto $t-1$ y $t+1$;
- Avanzado: Utiliza la velocidad instantánea, las variaciones de velocidad respecto a $t-5$, $t-3$, $t-2$, $t-1$, $t+1$, $t+2$, $t+3$ y $t+5$ multiplicadas por la velocidad instantánea al cuadrado, la velocidad instantánea al cubo, y las variaciones de altitud respecto $t-1$ y $t+1$.

En donde los hiperparámetros fueron el grado del polinomio y uso “10-fold crossvalidation”, se usó (10 particiones para medir el error en cada una de ellos).

Utilizó varias técnicas de Machine Learning y obtuvo resultados mejores en el escenario avanzado, para las diferentes técnicas. El investigador verificó que “Support Vector Machine” captura las no linealidades en una medida menor que “Random Forest” y “Extreme Gradient Boosting”. En el escenario medio, la técnica que no se ajustó sola y que necesitó variables de entrada ajustadas fue el modelo de regresión lineal.

En su análisis de comparación indicó que cuanto mayor diferencia de error hay entre los errores de entrenamiento y los errores de prueba, el modelo se sobre ajusta. Un sobre ajuste (“overfitting”) tiene problemas para predecir resultados aceptables con datos de entrada nuevos. El estudio lo realizó en R-studio.

El trabajo de Martín publicado en 2019 respecto al tema de contaminación indicó que según datos de la OMS en 2016 que el 91% de población vivía en lugares donde no se respetaba las directrices de la OMS [20] y en la

actualidad más del 99% de la población vive en zonas donde la contaminación atmosférica está por encima de las directrices de calidad del aire y 4,2 millones de muertes se atribuyen a la contaminación del aire ambiente cada año [21].

Motivado por esta situación, Martín utilizó datos proporcionados por INSIA, con información de emisiones, variables cinemáticas, ambientales y condiciones del motor tomados en 2007, 2008 de equipos montados en buses de la Empresa Municipal de Transporte de Madrid. Los datos fueron agrupados según el combustible y carga: B100 plena carga, B100 media carga, B100 vacío, Gasóleo media carga y Gasóleo vacío. En este trabajo Martín utilizó técnicas de Machine Learning (redes neuronales recurrentes) para modelar y predecir emisiones de CO₂ a partir de los datos sobre variables cinemáticas y de condiciones ambientales. Trabajó con 73852 observaciones y 8 variables. Indica, además, que trabajos anteriores mostraban errores de estimación altos y menciona el trabajo de Nova publicado en 2018 [22] y por eso propuso el uso de redes neuronales recurrentes para mejorar los resultados. Trabajó con series de tiempo, variables de velocidad, aceleración y sobre aceleración y respecto a variables ambientales concluyó que las que proporcionan mejores resultados fueron la temperatura y la humedad. También utilizó las variables de carga como categóricas para analizarlas conjuntamente y generar un modelo único en vez de modelos por separado. Para el pre procesamiento de datos tomo en cuenta valores erróneos, no disponibles ó NA, transformaciones lineales y no lineales y división de datos de entrenamiento (60%), validación (30%) y test (10%). Finalmente, realizó dos modelos uno de una sola red y otro de 5 redes, una capa oculta con 9

neuronas las que mejores resultados dan. Concluyó además que aún cuando la media de error para ambos modelos es alrededor de 60%, la mayoría de los resultados de predicciones tienen un error cerca del 15%, pudiéndose atribuir estos errores a datos con valores atípicos. El entorno utilizado fue R studio y redes neuronales recurrentes.

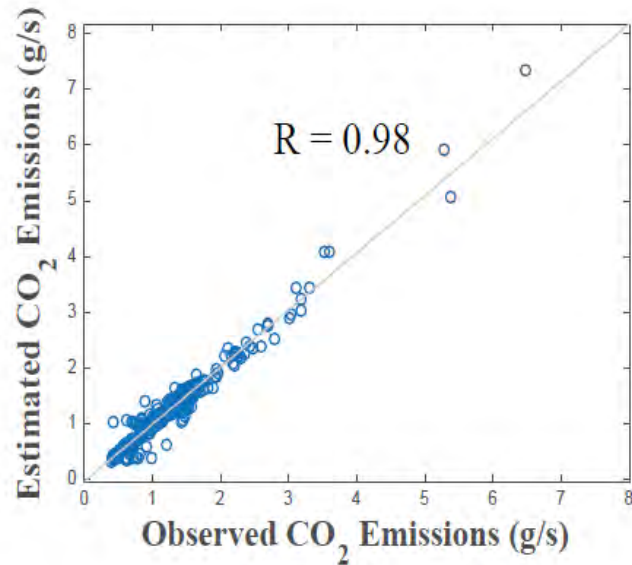
La publicación en la revista “Environment Pollut Climate Change” por Li, Qiao y Yu en 2016 [23], (DOI: 10.4172/2573-458X.1000106), muestra cómo se realizan modelos para estimar las emisiones de CO₂ y otros gases de escape para vehículos ligeros con técnicas de Machine Learning. Estos modelos se enfocan en las emisiones durante la marcha en vacío (ralentí). los autores indican que, aunque las emisiones durante las marchas en vacío son cortas, tienen un impacto importante en las emisiones. Se dan dos casos, ralentí discrecional y no discrecional, el primero se da a voluntad del conductor y el segundo se da por situaciones de tráfico, señales de tráfico entre otros. También se indica que hay modelos para modelar emisiones y uno de los más populares es MOVES (“Motor Vehicle Emission Simulator”) o el modelo europeo COPERT (Computer Program to Calculate Emissions from Road Transport). Estos modelos son macroscópicos y son muy útiles para escalas regionales; sin embargo, también es posible utilizar estos modelos en escala microscópicas y para ello se utiliza tasas de emisiones en ciclos de manejo (modelos microscópicos) pero los resultados obtenidos de esta manera terminan con sobre ajustes con resultados no esperados. Los autores realizaron el modelo y para ello utilizaron un vehículo ligero marca Subaru, modelo Forester 2004, de 2.5 litros, 165 hp a 5699 rpm y torque de 225Nm a 4000 rpm, con motor a gasolina. El modelo trata de pronosticar las emisiones

en ralentí en condiciones de conducción real y se realiza en varios tipos de carreteras, para esta prueba se usó un sistema PEMS (para toma de datos de CO₂, CO y HC) y los datos de motor rpm, consumo de aire y temperatura en información ambiental tomado del OBD2. Los autores aplicaron 5 técnicas de Machine Learning.

Para este estudio se registraron 8,637 pares de datos durante ralentí, registrados en 221 momentos. Los errores relativos para el CO₂ fueron: "K-Nearest Neighbors" (KNN) 2%, "Bagged Decision Tree" (BBDT) 1%, Redes Neuronales 3%, "CHI-squared Automatic Interaction Detection" (CHAID) 50% y "Support Vector Machine" (SVM) 2%.

Los resultados muestran que obtiene mejores modelos de ajuste para. Para emisiones de CO₂ en validación se obtiene $R=0.98$ y para test $R=0.98$. En el caso de KNN, para validación se obtiene un $R=0.94$ y para test un $R=0.93$. el RMSE validación para CO₂ (g/s) fue 0.19 y para test 0.21. En el caso de BBDT, para validación 0.10 y para test 0.08.

En la Figura 12 los valores de $R=0.98$ muestran una alta correlación para fase test en el caso de BBDT. Los valores encontrados para el caso de CO tienen un error de 10% y los autores concluyen que el modelo BBDT puede ser embebido dentro de un smartphone mediante el desarrollo de una aplicación que indique mediante una señal de las emisiones de en ralentí durante una congestión.



(a) CO₂ testing results

Figura 12 - Ajuste de correlación fase Test BBDT tomado de referencia [23].(DOI: 10.4172/2573-458X.1000106).

El trabajo publicado por Rosales en 2021 para la revista Tecno Humanismo [24] menciona el peligro asociado a las emisiones de CO₂ en el calentamiento global y también en su impacto en nuestro país, con inundaciones en costa, friajes en la sierra y selva. El estudio de Rosales utiliza datos del Banco Mundial y los modelos usados son ARIMA y Redes Neuronales. En su estudio utiliza 59 observaciones de las emisiones de CO₂ en el periodo 1960 al 2018, de los cuales tomó como datos de entrenamiento los datos hasta el 2014 y los datos para prueba los correspondientes a los años 2015 al 2018. Finalmente hace una comparación de los resultados y obtuvo que para Redes Neuronales un RMSE 707.13 para entrenamiento y RMSE de 1125.82 para prueba, mientras que en el mejor de los casos ARIMA obtuvo un RMSE de 1590.56

para entrenamiento y un RMSE de 4223.73 para datos de prueba. Menciona además que Redes Neuronales trabaja mejor con serie de datos con no linealidades. La estructura de la Red Neuronal empleada fue de 5 entradas, 10 neuronas en la capa oculta, conectado a una capa de salida.

Herrera [25] aplicó Redes Neuronales para predecir el índice de contaminación por CO₂ en la ciudad de Santo Domingo, en Ecuador. Para este estudio la toma de datos la realizó mediante un dispositivo electrónico basado en Arduino, acoplado con un sensor MG811 (usado para detección de CO₂). Este sensor envió señales analógicas al Arduino y, este, lo almacenó como datos numéricos, haciendo el registro de hora y fecha. Los datos registrados fueron contrastados con el medidor “Extech CO250”.

Para el análisis de datos utilizó el Software Matlab. Las predicciones las realizó para diferentes periodos de tiempo, y para a evaluar su modelo utilizó el Error Cuadrático Medio (MSE). Previo al análisis, Herrera realizó una depuración de los datos adquiridos, para eliminar datos atípicos. Los resultados de su predicción por día tienen valores promedio diario de 599 ppm (real) y 597.30 ppm (estimado) en el ambiente, con un error promedio de 1.27 ppm. Para el modelado por semana, los valores reales tienen un promedio de 598.5 ppm y la predicción tiene 596.9 ppm, con un error promedio de 1.60 ppm. La estructura de la Red Neuronal utilizada por Herrera tiene, cuatro entradas, una salida, y diez neuronas en la capa oculta.

Ortiz y Ceballos [26], realizaron un trabajo con el uso de inteligencia artificial en el análisis de datos para pronosticar costos en una empresa de transportes. Este trabajo inició haciendo una correlación de variables y se

eligieron las que consideraron importantes. Los datos considerados fueron: costo de combustible utilizado; tiempo de inicio de viaje hasta el encendido del motor; tiempo de conducción; cantidad de combustible consumido en cada viaje y la distancia recorrida en el viaje. Luego prepararon los datos (normalizaron) y eliminaron los datos atípicos. Los autores utilizaron Regresión lineal, "Random Forest Regression", "Gradient Boosted Trees" y Redes Neuronales. Trabajaron con un total de 8350000 registros, menos unos 84305 que eliminaron por considerarlos atípicos. Para evaluar su modelo usaron MSE, RMSE MAE y R^2 . De los resultados para las diferentes técnicas, los autores indicaron que, para los datos de entrenamiento, Regresión Lineal obtuvo un R^2 de 0.7132 (el más alto de todas las técnicas), MSE de 0.000, RMSE de 0.0015, y MAE de 0.0004. Las técnicas "Random Forest Regression" y "Gradient Boosted Trees obtuvieron un R^2 de 0.6862 y 0.6909, respectivamente. Redes Neuronales obtuvo R^2 de 0.0126 (muy bajo), con MSE de 0.0732, RMSE de 0.2706 y MAE de 0.1479. Con estos resultados, los autores descartaron el modelo de Redes Neuronales para los datos de prueba, por lo que solo mostraron los resultados para los tres modelos restantes. Para los datos de prueba, los valores obtenidos fueron: para Regresión Lineal, R^2 de 0.8157, MSE de 0.0461, RMSE 0.2146 y MAE de 0.2044; para "Gradient Boosted Trees", MSE de 0.000, RMSE de 0.0015, MAE de 0.0003 y R^2 de 0.6909; para "Random Forest Regressor", MAE de 0.0813, RMSE de 0.2850, MAE de 0.2250 y R^2 de 0.6750. Los resultados de este estudio confirmaron que el modelo de Regresión Lineal presentó un mejor ajuste, y también para datos de prueba.

Concluyendo con este capítulo, reafirmamos que en el presente trabajo de tesis se utilizarán las metodologías Regresión Lineal Múltiple, Máquina de Soporte Vectorial (SVM), como lo empleado por Madziel [18] (DOI: 10.3390/en15010142), y Caro [19], así como Redes Neuronales [23] (DOI: 10.4172/2573-458X.1000106). Los datos para la realización de los modelos serán de tipo experimental, para lo cual tomarán parámetros instantáneos de funcionamiento del motor (por tratarse de modelo de escala microscópica) obtenidos como parte del trabajo desarrollado.



CAPÍTULO 3

DESARROLLO

Este trabajo presenta un enfoque cuantitativo y de tipo regresión que desarrolla un modelo mediante diferentes técnicas de Machine Learning de tipo supervisado para predecir el consumo de gasolina y emisión de CO₂ con los datos obtenidos por experimentación de un motor de 130 HP, instalado en un banco de pruebas. El estudio tiene procesos secuenciales para su desarrollo.

Población y datos: corresponden a los parámetros del funcionamiento del motor a gasolina de 130 HP y la frecuencia de recolección de datos, uno por segundo. Los pasos de la experimentación se pueden resumir del siguiente modo:

1. Recolección de datos tomados del banco de pruebas almacenados en dos tipos: unos serán variables predictoras y otras variables objetivo;
2. Análisis de los datos, división, estandarización, elección del algoritmo para el modelado (elección del algoritmo) y predicción;
3. Elaboración del reporte de resultados.

La investigación, en esta primera etapa, fue del tipo experimental.

3.1. Equipos empleados

- Banco de pruebas de motor Hyundai Elantra 130 HP;
- Dinamómetro;
- Medidor de aire de flujo laminar;
- Ventilador industrial (para refrigeración del motor);
- Medidor de combustible tipo Coriolis;
- Transmisión de presión diferencial EJX110A;
- Tacómetro Monarch ACT-3X
- Registrador de datos electrónico;
- Analizador de gases de escape;
- Scanner con conexión Bluetooth;
- Equipo de cómputo de escritorio;
- Smartphone con aplicación “OBD fusion” (similar a aplicación “Torque”).

A continuación, se describen los equipos empleados:

El motor del banco de pruebas es Hyundai Elantra 1600cc - GAMMA, motor DOHC - MPI de 4 cilindros en línea, con capacidad 1600 cc, motor de 4 tiempos encendido por chispa y el sistema de distribución tiene doble árbol de levas en la culata (Double Overhead Camshaft), con 16 válvulas e inyección multipunto, con normativa de emisiones EURO 5. El tipo de combustible indicado para este motor es gasolina sin plomo. Potencia máxima 130 HP, par máximo 157 Nm. Tiene una relación de compresión 10.5 a 1, con diámetro de cilindro 77 mm y carrera 85.44 mm

Los datos del motor de combustión interna Hyundai 130 HP se muestran en la Tabla 1 y en Anexo B.

Tabla 1 - Características del motor Hyundai laboratorio PUCP

TIPO	1.6 MPI – GAMMA DOHC
NÚMERO DE CILINDROS	4
NÚMERO DE VÁLVULAS	16
POTENCIA MÁXIMA (HP/RPM)	130 / 6,300
TORQUE MÁXIMO (KG-M/RPM)	15.7 / 4,850
CILINDRADA (CC.)	1591
TIPO DE COMBUSTIBLE	GASOLINA SIN PLOMO

Para este trabajo se utilizó gasolina Premium de 95 octanos como mínimo y con un contenido de azufre menor a 50 ppm, ver ficha técnica en Anexo C.

Las principales vistas del módulo de control del banco de pruebas se presentan en las Figuras 13 (módulo de control del dinamómetro) y Figura 14 (panel de control para el motor Hyundai 1600 cc). El panel de control del motor cuenta con un interruptor para encendido, reloj indicador de RPM, medidores de presión de aceite, medidores de temperatura de motor y amperímetro.



Figura 13 - Módulo de control del Dinamómetro.



Figura 14 - Módulo de control del motor Hyundai 1600cc.

El dinamómetro utilizado es de marca Zoller, modelo A350 (ver equipo de color verde, en la Figura 15). Posee dos rangos de medición, de 0-150 Nm y 0-300 Nm, con una velocidad máxima de 6500rpm controlados por un panel ajusta el torque aplicado al motor. La Figura 15 muestra el eje cardan que conecta el motor con el dinamómetro.

Con dos rangos de medición 0-150 Nm y 0-300 Nm y con una velocidad máxima de 6500rpm controlados por un panel que controla rpm y torque. La marca del Dinamómetro es Zoller. Modelo (A350). La ilustración muestra el eje cardan que conecta el motor con el dinamómetro.



Figura 15 - Dinamómetro conectado al motor con un eje cardan

El Medidor de aire de flujo laminar (véase Figura 16) fue empleado para contrastar los valores estimados por el software de la ECU del motor y vía puerto OBDII mediante scanner y aplicación “OBD fusión”.



Figura 16 - Medidor de aire de Flujo laminar.

El uso de un ventilador axial (ver Figura 17) fue necesario debido a que el motor eleva su temperatura durante la prueba (su sistema de refrigeración está estimado para tener el flujo de aire debido al movimiento del automóvil) y su uso ayuda a no tener problemas mecánicos debido al recalentamiento.



Figura 17 - Ventilador axial

Medidor Coriolis Micro Motion Elite CMF010M, señal de salida de 4-20 mA, temperatura de proceso de -240°C a 350°C y presión de operación máxima 125 barg, presión de la densidad de líquido $\pm 0.2\text{Kg/m}^3$. Con certificación UL



Figura 18 - Transmisor y medidor de flujo tipo coriolis.

Modelo Micro Motion 2700 (Transmisor de micro movimiento), con rango de trabajo de 5 a 95 % HR en un rango de temperatura de -40 a +60 °C, fuente de 18 a 100 VCC - 85 a 265 VCA.

Transmisor de presión diferencial YOKOGAWA modelo EJX110A (Figura 19) tiene las siguientes características: 0.04% de precisión (opción disponible de precisión de 0.025 %), $\pm 0.1\%$ de Estabilidad; 90 milisegundos de tiempo de respuesta; 3,600 psi MWP; Certificación Exida y TUV SIL2 / SIL3 y ajuste de parámetro local.



Figura 19 - EJX110A Transmisor de Presión Diferencial.

Tacómetro ACT-3X (ver Figura 20), panel tacómetro de 5 - 999999 RPM con certificado de calibración trazable del panel frontal programable uno o múltiples pulsos por revolución. Límite de tensión de salida analógica 0 – 5 Vcc, 4 – 20 mA y máximo de línea de alimentación 220/230/240Vca, 50/60 Hz. 12 o 24 Vcc.



Figura 20 - Tacómetro Monarch, modelo ACT-3X

El registrador YOKOGAWA de tipo portátil (Figura 21), con registro electrónico y almacenaje externo mediante memoria SD se utilizó para el registro de los equipos y sensores instalados en el motor y banco de pruebas durante la segunda etapa de pruebas preliminares en adelante.



Figura 21 - Registrador YOKOGAWA.

Así mismo la recolección de datos del ECU del motor se realizó mediante el puerto OBDII.

Analizador de los gases de escape vehicular marca Infrared Industries, modelo FGA4500 (ver Figura 22), clase de exactitud OIML Clase 1, con resolución CO: (0 a 10,00) %volumen; CO₂: (0 a 20) %volumen; HC: (0 a 10000) ppm; O₂: (0 a 25) %volumen. Ver certificado de calibración en Anexo D.



Figura 22 - Analizador de gases de escape vehicular, marca Infrared Industries, modelo FGA4500.

El scanner OBDII tiene las siguientes características: marca OBD Link; conexión bluetooth. Este scanner realiza las siguientes lecturas: DTC genéricos, lectura de parámetros genéricos en vivo, como muestra la cobertura específica del fabricante [27]. Alimentación mediante conexión al vehículo, dimensiones 2.01 x 0.91 x 1.81 pulgadas, sistema operativo iOS, Android y Windows.

Para el registro de datos proporcionados por el scanner OBDII se utilizó un teléfono celular, marca LG, modelo LG K50, sistema operativo Android Pie 9, con memoria RAM de 3 GB y memoria interna de 32 GB. La aplicación instalada en el teléfono celular para conexión con el scanner fue "OBD fusion"

El equipo de cómputo utilizado tiene las siguientes características: Procesador Intel(R) Core (TM) i5-10400F CPU @ 2.90GHz, con RAM instalada de 16.0 GB y sistema operativo de 64 bits.

3.2. Etapas de la experimentación

Primera etapa de la experimentación:

En la Figura 23 se presenta el diagrama de flujo de la primera etapa de la experimentación, luego se describe cada una de estas.

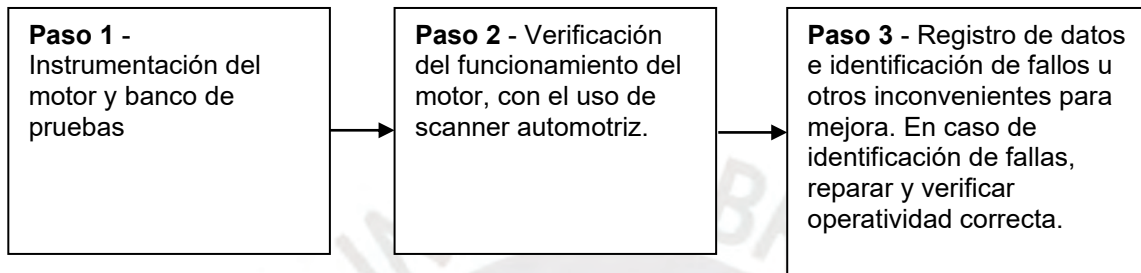


Figura 23 - Diagrama de flujo de pasos a seguir en pruebas de verificación.

Para medir el flujo de aire admitido al motor, se instaló un medidor de flujo del tipo placa de orificio, instalado en la entrada de un tanque pulmón (componente de color azul en la Figura 24). La conexión entre la manguera del medidor de flujo fue mediante conector flexible y niple de acero inoxidable, de 2.5" de diámetro exterior.



Figura 24 - Accesorios de conexión y medidor del flujo de aire de admisión al motor.



Figura 25 - Vista General del medidor de flujo de aire y banco de prueba del motor.

Para registrar la variación de presión diferencial del aire a través de la placa de orificio, se instaló un medidor de la marca Dwyer 477 (ver Figura 26).



Figura 26 - Medidor de presión diferencial Dwyer 477

El acoplamiento entre el dinamómetro y el motor fue mediante un eje cardán, protegido por una carcasa para evitar accidentes (véase Figura 27).



Figura 27 - Detalle de conexión entre motor y dinamómetro mediante eje cardan.

El dinamómetro tiene dos rangos de trabajo que debe seleccionarse previo a las pruebas. En el presente estudio, el rango de trabajo fue de 0 a 150 Nm, debido al torque del motor bajo estudio (ver Figura 28).



Figura 28 - Dinamómetro en el rango de inferior con un máximo de 150 Nm.

Para medir el consumo instantáneo de gasolina, inicialmente se planteó usar una balanza, pero debido que se requiere un registro en el tiempo se optó por un medidor de tipo coriolis con un transmisor de datos. Debido a que el tanque diario de combustible se ubica a una distancia alejada del motor, fue necesario usar conectores para instalar en tanque auxiliar entre el motor y tanque diario, para lo cual se adquirió los materiales respectivos (ver Figura 29).

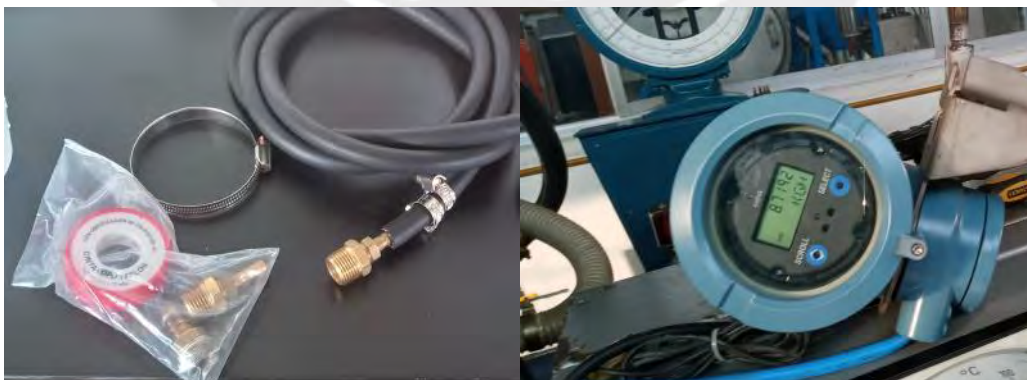


Figura 29 - Materiales usados y transmisor conectado a un medidor coriolis (lectura en kg/h).

El sistema de control de temperatura del dinamómetro e indicador visual de RPM se muestran en el tablero del banco de pruebas (ver Figura 30). Durante las pruebas, se verificó que no se supere el valor máximo de 30°C.

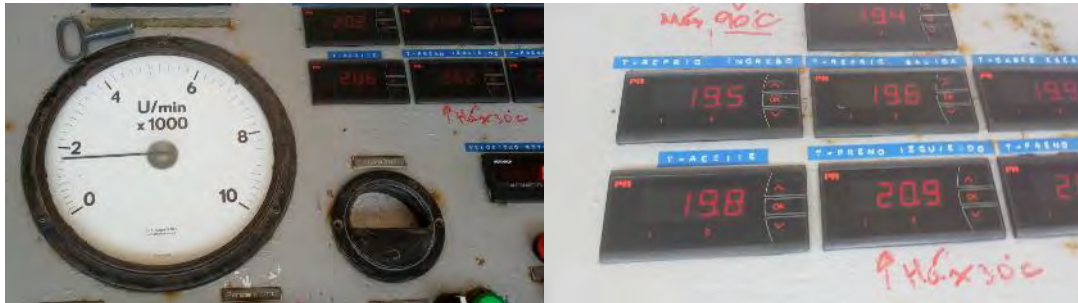


Figura 30 - Visor de RPM y control de temperatura asociado al dinamómetro.

La temperatura de funcionamiento del motor se mantiene, por un lado, por el sistema de enfriamiento del mismo motor y, por otro lado, mediante el flujo de aire forzado por un ventilador de la sala de pruebas (ver Figura 31).



Figura 31 - Ventilador para regular la temperatura del motor en banco de pruebas.

Para la verificación del normal funcionamiento se procedió a conectar un scanner al módulo de control del motor, por medio del conector OBDII. Este

scanner envía datos a un teléfono celular mediante conexión *bluetooth* (ver Figura 32).



Figura 32 - Dispositivo OBDII conectado a telefono celular vía aplicativo para captura de datos en tiempo real.

Instalada toda la instrumentación y equipos de registro de datos, se procedió a verificar el funcionamiento normal del motor y registradores de señal. Para ello, se puso en marcha el motor, tomando en cuenta que debe estar encendido sin carga hasta que alcance su temperatura de operación. Se verificó que el panel de control muestre la señal de “Check Engine” encendida, utilizando el scanner instalado. En caso de alguna falla, se procedió a buscar el código correspondiente, corrigió (si era necesario) y luego se borró el código (de fallas históricas); finalmente, se hizo un nuevo examen para verificar si presentaba códigos nuevos.

Inicialmente, las pruebas se definieron para torques variando de 30 a 94 Nm, en intervalos de 10 Nm; cada uno de los cuales fue evaluado en el intervalo de régimen de giro de motor de 1000 a 4000 RPM (véase Tabla 2).

Tabla 2 - Valores de torque (Nm) y régimen de giro (RPM) inicialmente previstos.

Ítem	Torque (Nm)	Rango replanteado de RPM	Observaciones
1	94	1700-3500	Sin problema de registro de datos
2	90	1600-4000	Sin problema de registro de datos
3	80	1500-4000	Sin problema de registro de datos
4	70	1400-3500	Sin problema de registro de datos
5	60	1000-3000	Valor registrado de valor 0 para en el medidor de presión diferencial (Dwyer 477)
6	50	700-2500	Valor registrado de valor 0 para en el medidor de presión diferencial (Dwyer 477)
7	40	600-2000	Valor registrado de valor 0 para en el medidor de presión diferencial (Dwyer 477)
8	30	600-1500	Valor registrado de valor 0 para en el medidor de presión diferencial (Dwyer 477)

Sin embargo, durante las pruebas preliminares se verificó dificultades del control dinamométrico para mantener fijo algunas configuraciones de torque x

RPM. Debido a esto, fue necesario replantear el rango de valores de torque y RPM (ver Tabla 3).

Tabla 3 - Rango de torque (Nm.) y rango de giro de motor (RPM) replanteados.

Ítem	Rango de torque planteado (Nm.)	Rango de RPM planteado
1	92-80	1600-3500
2	90-72	1500-4000
3	82-47	1500-3500
4	74-26	1500-3500
5	57-10	1000-3000
6	100-93	2000-3000
7	107-93	2600-3500
8	72-28	1000-2500

Luego de finalizada la prueba preliminar se identificó otros inconvenientes: El primero fue debido al ruido generado por el motor hacia el exterior (salida de los gases de escape) y el segundo fue que el mecanismo de regulación del pedal del acelerador tenía defectos en su diseño y era necesario realizar un nuevo mecanismo.

El ruido provocado por los gases de escape fue corregido con la instalación de un silenciador, luego del catalizador.

Para la fabricación del mecanismo para control del acelerador se decidió que debía ser considerando un actuador lineal, accionado por un motor paso a paso, NEMA 23. Para la elección del motor se hicieron mediciones de la fuerza necesaria para mover el pedal, obteniéndose el valor de 3.3 kgf (32.33 N), ver Figura 33.



Figura 33 - Medición de fuerza para mover pedal de acelerador.

Por lo tanto, el motor debería tener un torque aplicado a una polea de un centímetro de radio que genere una fuerza que sea superior a los 3.3 kgf. Un motor NEMA 23, de 24 V y 2.8 amperios, y 12.6 kgf.cm de torque (el más próximo comercialmente disponible), garantizó el requerimiento. El driver para este motor NEMA 23 (de 2.8 amperios) debería suministrar la corriente necesaria, por lo que seleccionamos un driver de 4 amperios y 24 V.

En las Figuras 34 y 35 se presentan el diseño y fabricación del soporte para actuador y controlador del pedal del acelerador del motor.

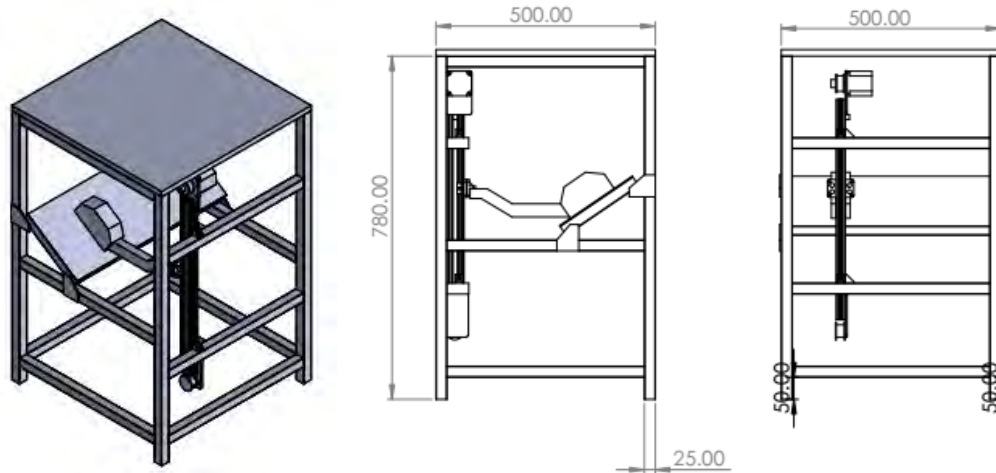


Figura 34 - Diseño de mesa para instalación de actuador lineal "V-slot".

Con la idea clara sobre el diseño se procede a realizar la fabricación en perfil cuadrado de 25mm de sección por 1.5mm de espesor (Figura 34).



Figura 35 - Proceso de fabricación y uso del soporte del actuador y controlador del pedal.

En la Figura 36 se aprecian los detalles del actuador lineal y soporte en el interior del panel de control del dinamómetro.



Figura 36 - Soporte instalado en el panel de control, con actuador lineal instalado (incluido motor NEMA 23).

En otro momento, también fue necesario el cambio del medidor de presión diferencial para medir el flujo de aire hacia el motor, junto a un registrador de datos, a causa de una mejor precisión en las lecturas. Se cambió el medidor de presión diferencial Dwyer 477 por el transmisor de presión diferencial YOKOGAWA, modelo EJX110A, también se instaló el registrador de datos YOKOGAWA que registra los datos en una memoria SD.



Figura 37 - Instalación de registrador y cambio de medidor diferencial de presión Dwyer 477 por YOCOGAWA modelo E110A.



Figura 38 - Captura de pantalla del registrador YOCOGAWA durante su configuración por técnico de laboratorio.

Para asegurar la cantidad y calidad de los datos se hizo una revisión de la estadística, que aseguré que la partición de datos, como son: entrenamiento, validación y prueba cubran todo el rango y tengas similares características; es decir, sean representativas del total (antes de la partición) y esto nuevamente se contrastó con los datos de la hoja de especificación del fabricante del

motor. Se excluyeron aquellos datos que están fuera del rango o presentaran mediciones nulas.

3.3. Implementación de los modelos de Machine Learning

Las modelos a realizar son dos, un modelo para predecir las emisiones de CO₂ y otro para predecir el consumo de combustible, las variables involucradas son:

Los criterios para la selección de las variables fueron los siguientes y se eligió por estar relacionado con la generación de CO₂.

Variables predictoras:

- Presión absoluta en el colector de admisión, en kPa.
- Temperatura del aire en la admisión, en °C;
- Régimen de giro del cigüeñal del motor, en RPM;
- Flujo másico de aire, en g/s;
- Torque efectivo del motor en N.m.

Variables objetivo:

- CO₂ g/s (modelo uno);
- Consumo de combustible kg/h (modelo dos).

El método requiere de la instrumentación necesaria y se realizó una prueba que facilitó el uso correcto de la instrumentación y también mediante un scanner conectada al puerto OBD II. Se tuvo en cuenta las unidades y una idea del orden de magnitud de los datos antes de realizar su registro.

3.4. Condición de repetibilidad

Dado que la prueba se realiza en un banco de prueba cumple con el requisito de comprobación de la repetibilidad en caso que sea necesaria repetir el experimento.

La frecuencia de recolección de datos es cada segundo y se recolectaron un total de 22370 registros, cada uno con cinco variables predictoras y dos variables objetivo.

El rango de los datos recolectados se muestra a continuación:

- Presión absoluta en el colector de admisión (kPa) : [21, 99]
- Temperatura en la entrada al múltiple de admisión (°C): [32, 47]
- Régimen de giro del cigüeñal del motor, en RPM : [618,5, 6270,8]
- Flujo de aire (g/s) : [0, 78.394]
- Torque efectivo del motor en N.m. : [17.300, 85.900]
- CO₂ (g/s) : [0, 10.400]
- Gasolina (kg/h) : [0, 12.416]

Se observa valores nulos para Flujo de aire, CO₂ y Gasolina. Estos serán revisados posteriormente en el preprocesamiento de datos.

3.5. Preprocesamiento de datos

Haciendo un análisis de datos nulos, encontramos que existen variables con valor 0 en flujo de aire, y lo mismo en flujo másico de CO₂. Se elimina el

registro que presenta valores nulos. El número de registros queda en 21 985 y el rango queda de la siguiente forma.

- Presión absoluta en el colector de admisión (kPa) : [21, 99]
- Temperatura en la entrada al múltiple de admisión (°C): [32, 47]
- Régimen de giro del cigüeñal del motor, en RPM : [618.5, 6270.8]
- Flujo de aire (g/s) : [1.467, 78.394]
- Torque efectivo del motor en N.m. : [17.3, 85.9]
- CO₂ (g/s) : [0.4, 10.4]
- Consumo de gasolina (kg/h) : [0.466, 12.416]

Una consideración importante, no tomada en cuenta, fue el tiempo de toma de datos, siendo más notorio en la toma de los datos del escape. Para modelos de emisión instantáneas las emisiones pueden ser relacionadas de forma correcta con el estado del motor si están correctamente localizadas en la escala del tiempo, los gases de escape al salir tienen un retraso respecto al estado actual del motor debido al recorrido hasta su salida [28], (DOI: 10.1016/j.atmosenv.2004.03.080).

A continuación, se muestra un diagrama de dispersión de las diferentes variables predictoras versus variable objetivo para las dos regresiones, la primera para el CO₂ en la Figura 39 y la segunda para el combustible en la Figura 40.

En el caso del CO₂, se observa que las variables predictoras presión en el múltiple de admisión, RPM, Aire y torque tienen una correlación positiva.

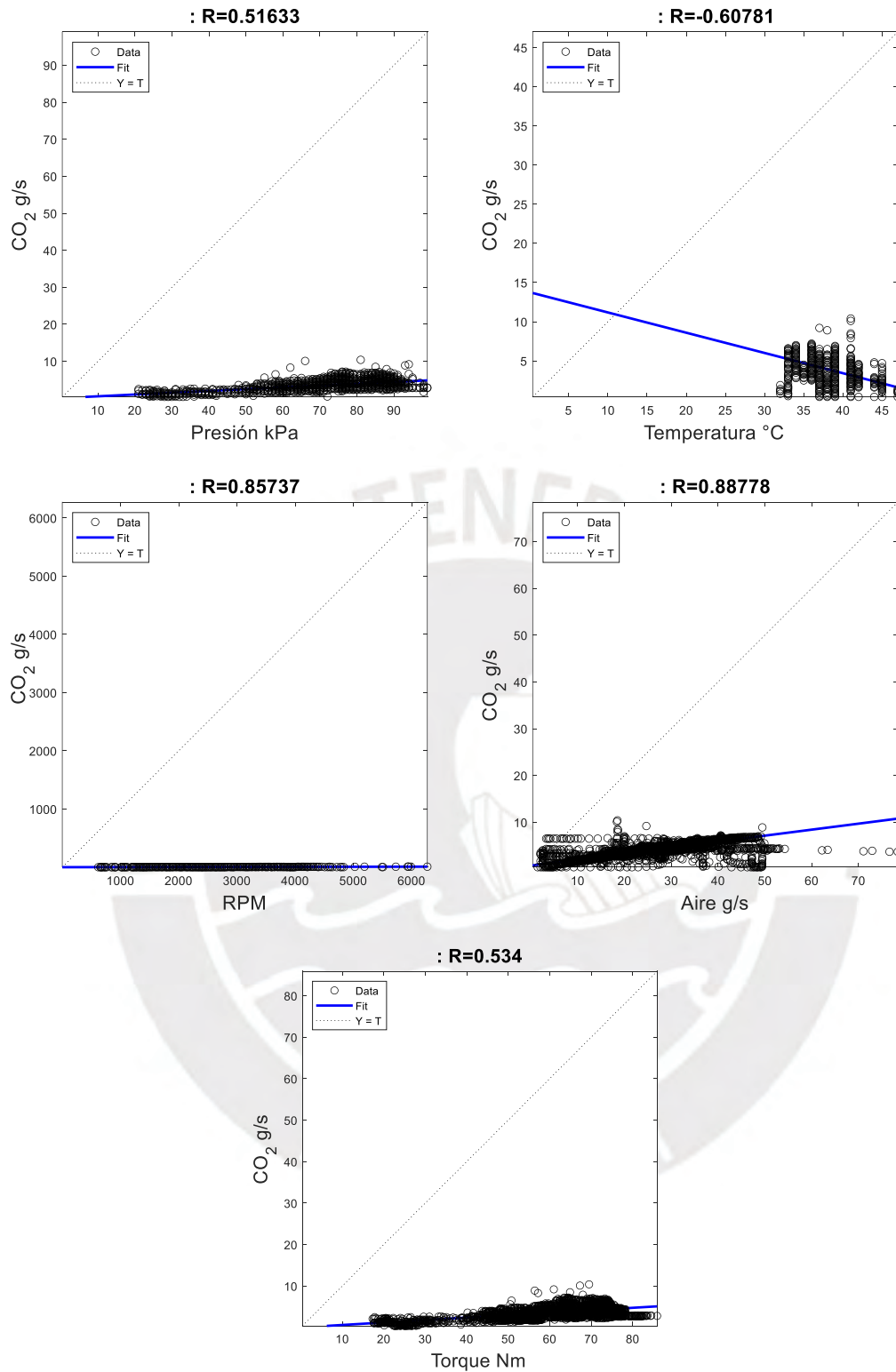


Figura 39 - Dispersión de diferentes variables Vs CO₂.

La Figura 40 muestra que la temperatura tiene una correlación negativa.

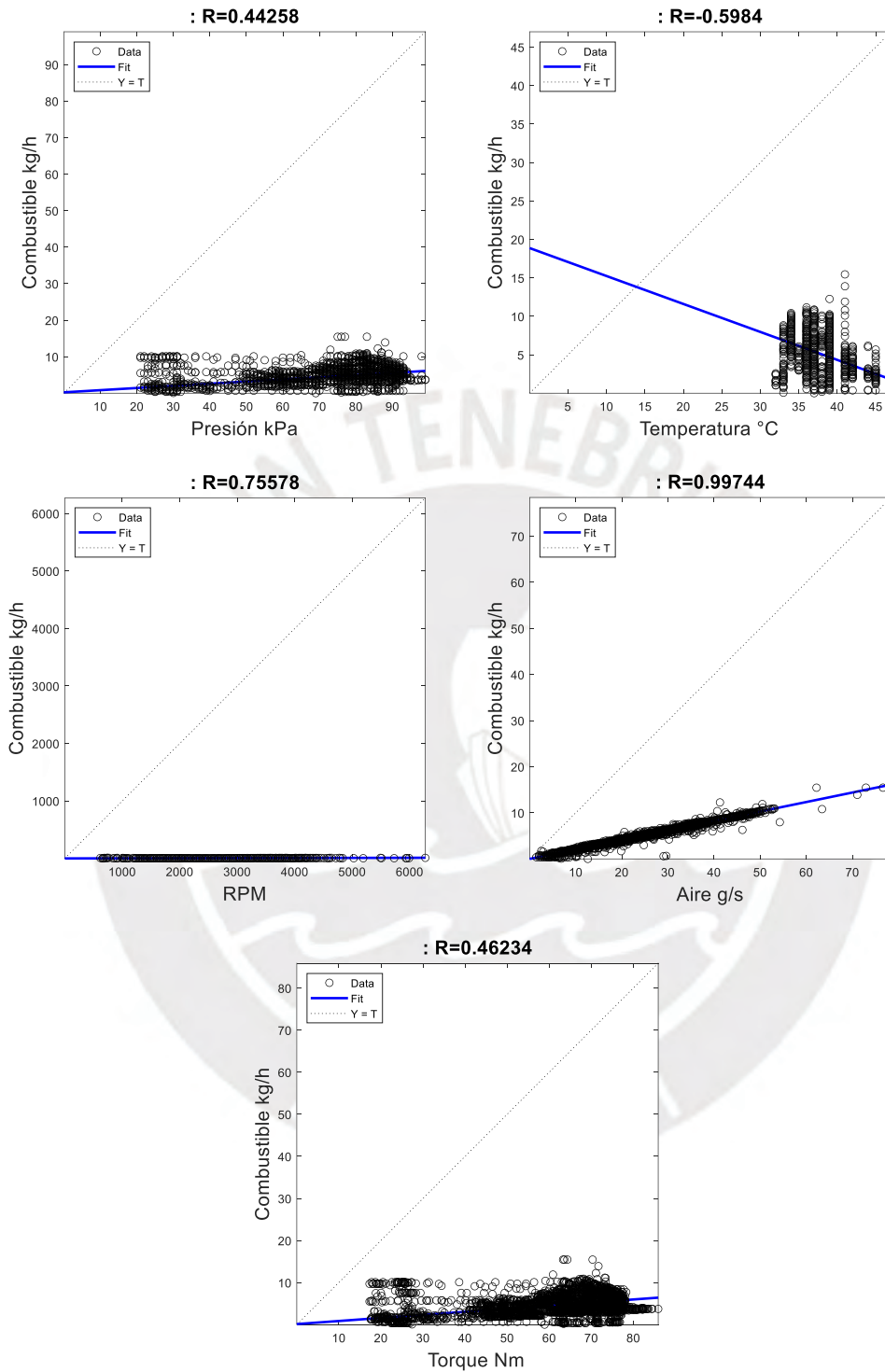


Figura 40 - Dispersión de diferentes variables vs consumo de gasolina. Elaboración propia.

Se puede ver los coeficientes de correlación lineal positiva para las variables Presión en la admisión, RPM, Flujo de aire en la admisión y torque, en el caso de la variable presión se observa baja correlación y para temperatura en el múltiple de admisión se observa correlación negativa (Ver Figura 39 y 40), pero conservaremos el dato para ver su impacto con otras herramientas como distancia de Cook.

3.6. División de datos

Se realizó una selección aleatoria datos de entrenamiento y prueba (90% y 10%) respectivamente. Con el fin de hacer la comparación entre las diferentes técnicas de Machine Learning.

La Figura 40 muestra los diagramas de caja o también llamado de bigotes [29] e indica que la selección aleatoria representa al rango completo de las variables y muestra además una media y medianas similares, tanto para entrenamiento (izquierda de la Figura 41) y prueba o test (derecha de la Figura 41).

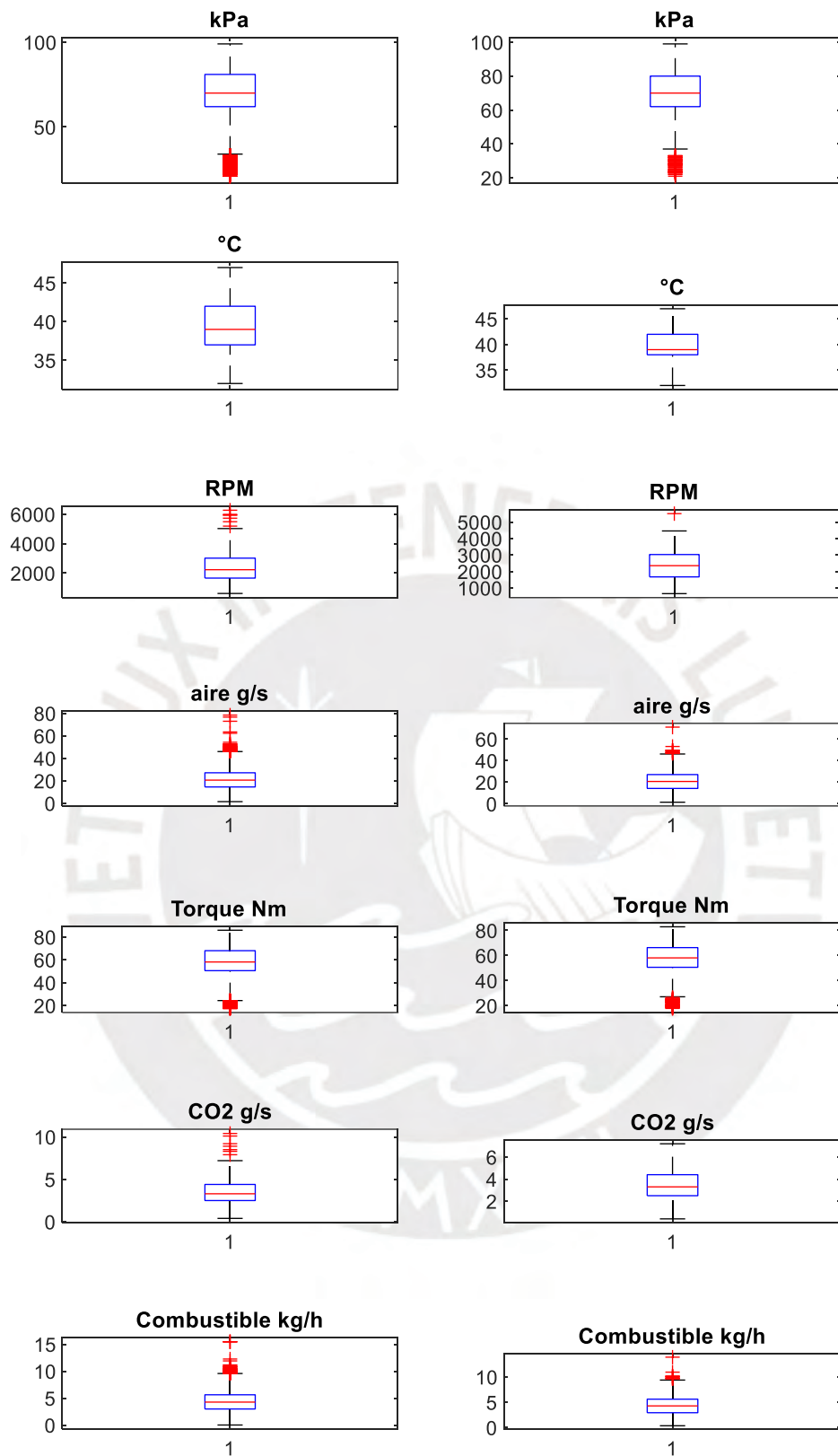


Figura 41 - Partición de datos: entrenamiento (izquierda), test (derecha).

Al realizar un histograma de las variables se observa claramente en la Figura 42, que la distribución es similar para los datos de entrenamiento y para los datos de test.

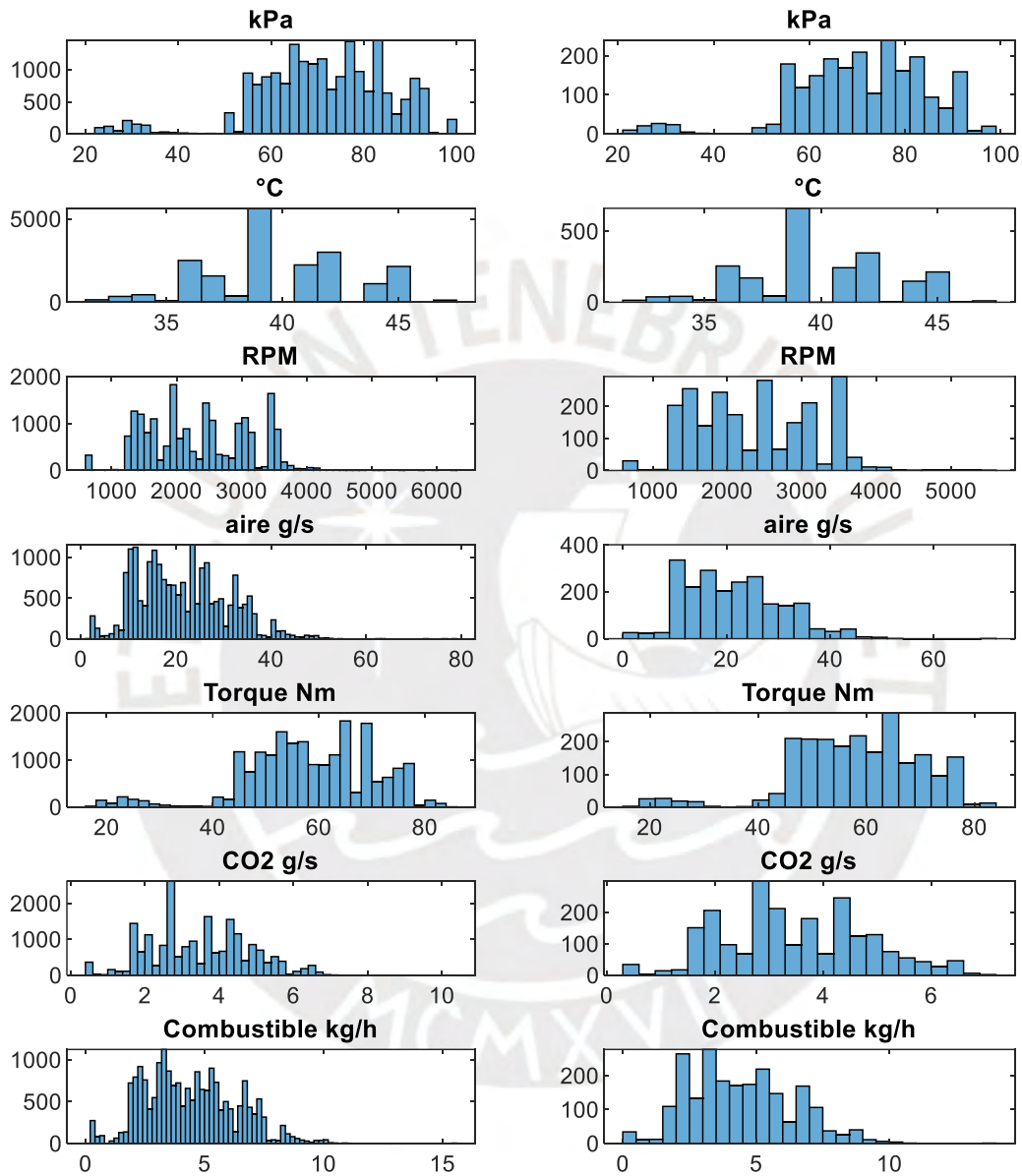


Figura 42 - Partición de datos: entrenamiento (izquierda), test (derecha).

También de la Figura 42 muestran asimetrías y el rango de los datos es similar para el conjunto de datos de entrenamiento como de test. En el caso

de datos como estos, de variable no discreta, no se puede elaborar la tabla de frecuencias y mas bien, se agrupa en intervalos de clase.

La primera división de datos nos da 90% para entrenamiento. Para el caso de Redes Neuronales realizamos una nueva separación en 3 partes: datos de entrenamiento, datos de validación y datos de prueba (75%, 15% y 15% respectivamente). Esta separación fue de forma aleatoria.

3.7. Estandarizar datos

Se transforman de modo que los nuevos datos tengan como media 0 y desviación estándar 1, esto favorece a que modelos de SVM y redes neuronales converjan. Este proceso se ejecuta a elección y no esta por defecto en el caso de Máquina de Soporte Vectorial, pero si en el caso de Redes neuronales.

3.8. Procesamiento de datos y generación de modelos

El enfoque del estudio es cuantitativo y contamos con variables objetivo de tipo no cualitativas por lo que usaremos técnicas de regresión. Los datos están clasificados en predictores y objetivo lo que nos permite utilizar técnica de Machine Learning de tipo supervisado, dentro los cuales se evaluarán:

- Regresión Lineal múltiple;
- Máquina de Soporte Vectorial;
- Redes Neuronales.

3.8.1. Regresión Lineal Múltiple

En la Tabla 4 se muestran los arreglos (modelo) de Machine Learning de regresión lineal utilizados en el presente trabajo. Se analizarán cinco modelos.

Tabla 4 - Modelos aplicados de Regresión lineal múltiple, tomados de Matlab [30] .

Machine learning	Modelo	Detalle en notación Wilkinson
mdl	linear	$y \sim 1 + x_1 + x_2 + x_3 + x_4 + x_5$
	interactions	$y \sim$ [Formula lineal con 16 términos y con 5 predictores]
	purequadratic	$y \sim 1 + x_1 + x_2 + x_3 + x_4 + x_5 + x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2$
	quadratic	$y \sim 1 + x_1x_2 + x_1x_3 + x_1x_4 + x_1x_5 + x_2x_3 + x_2x_4 + x_2x_5 + x_3x_4 + x_3x_5 + x_4x_5 + x_1^2 + x_2^2 + x_3^2 + x_4^2 + x_5^2$
	poly11222	$y \sim$ [Formula lineal con 19 términos y con 5 predictores]

Los términos de variables predictoras que intervienen están descritos en notación Wilkinson [31]; por ejemplo, el modelo **linear** tiene notación de operadores: $y \sim 1 + x_1 + x_2 + x_3 + x_4 + x_5$. El número 1 indica que el modelo lineal es de un término interceptor y cinco variables, todas de grado 1, es decir el modelo será de la forma $y_i = \beta_0 + \beta_1x_{i1} + \beta_2x_{i2} + \beta_3x_{i3} + \beta_4x_{i4} + \beta_5x_{i5} + \varepsilon_i$. El modelo **interactions**, contiene los términos predictores y sus combinaciones lineales incluidos el término predictor (por defecto se incluye y para excluirlo colocar -1). El modelo **purequadratic** incluye términos de variables predictores de orden 1, combinaciones de predictores y predictores de grado 2 y este también contiene el término interceptor. El modelo **quadratic** contiene combinación de predictores y términos de variables predictoras de grado 2 incluido el interceptor. En el caso de modelo

polyijk, ajustamos valores para los dos predictores a 1 y para los restantes a 2, esto debido a que los valores de los dos primeros predictores corresponden a presión y temperatura, los cuales no tiene correlación lineal.

Los datos se dividen en dos: 90% para la creación del modelo y 10% para el test.

Ver diagrama de flujo para modelo de Regresión Lineal Múltiple en Figura 43.



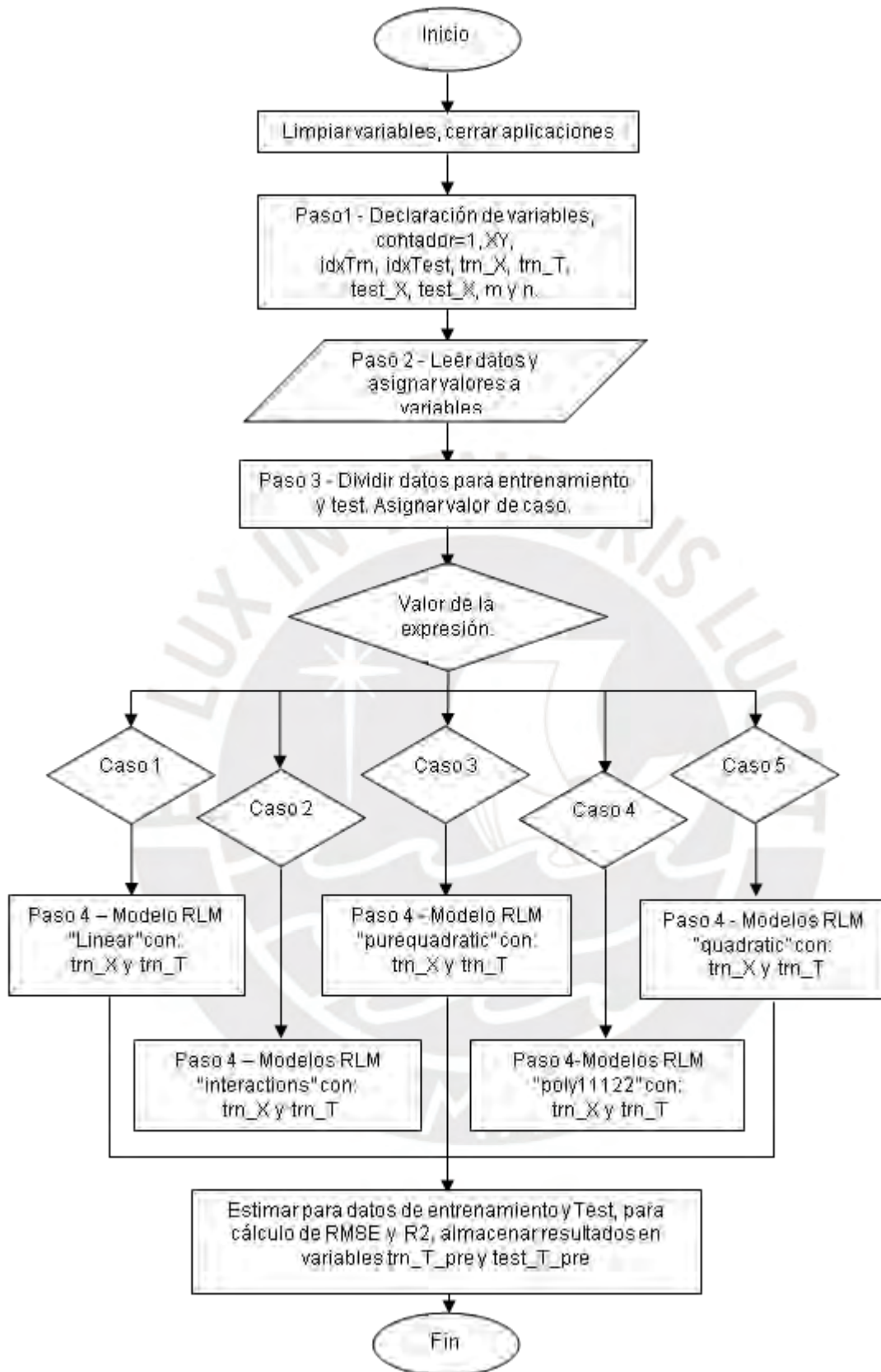


Figura 43 - Diagrama de flujo para algoritmo de Regresión lineal múltiple.

3.8.2. Máquina de soporte Vectorial (SVM)

Se analiza cuatro escenarios:

- Con datos no estandarizados;
- Usando datos estandarizados;
- Con datos estandarizados y función Kernel polinomial de segundo orden;
- Con datos estandarizados y función Kernel Gaussiana.

Se verifica la convergencia con el comando *Mdl.ConvergeInfo.Converged* y el resultado se puede asignar a una variable a la que la función devuelve el valor 0 ó 1. Un valor 0 indica que el procedimiento no converge y un valor 1 significa que converge. Nuevamente, se trabajó con el 90% para entrenamiento y 10% para test.

A continuación, se muestra el diagrama de flujo del modelo Máquina de Soporte Vectorial (ver Figura 44)

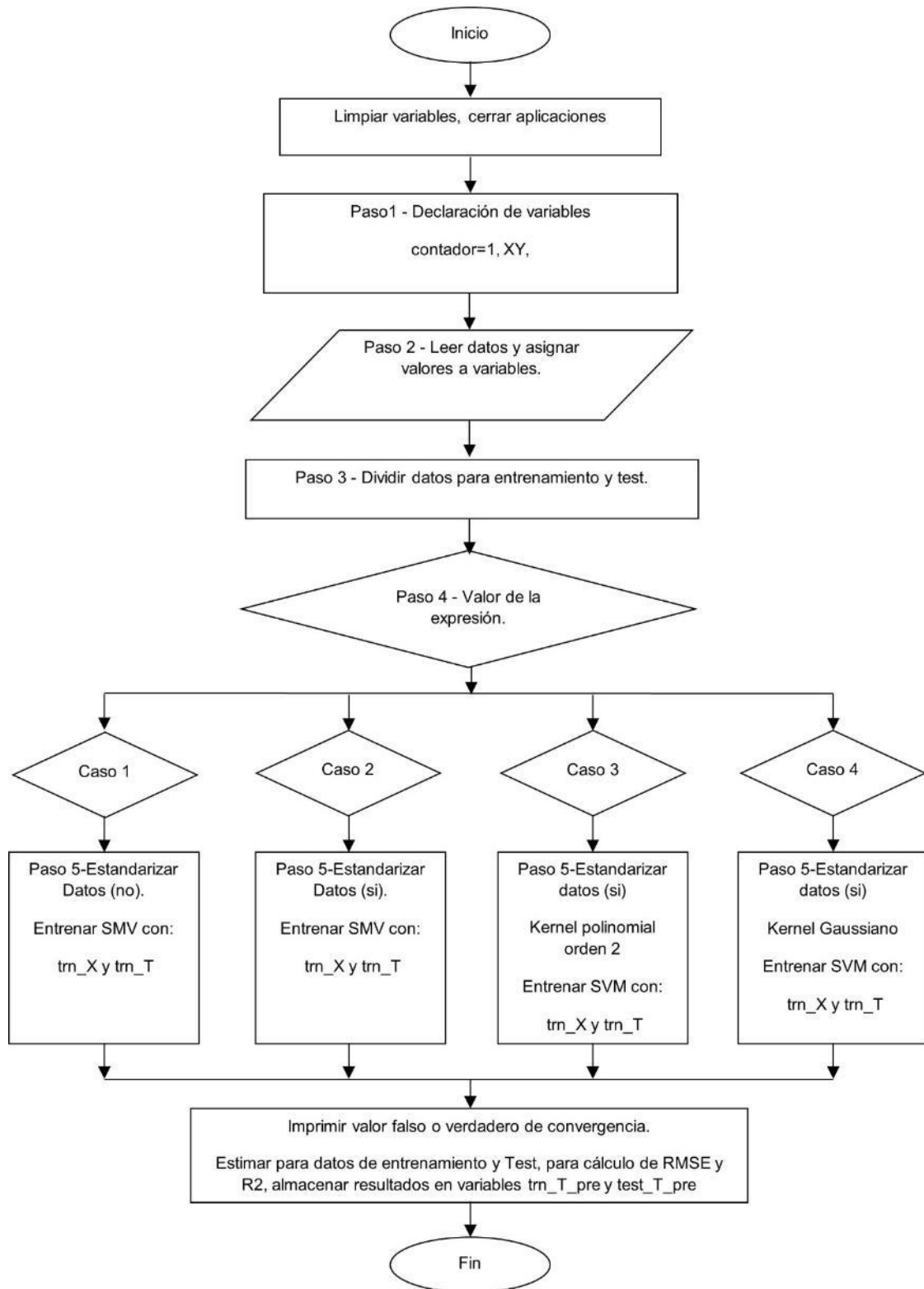


Figura 44 - Diagrama de flujo para algoritmo de Máquina de Soporte Vectorial para regresión.

3.8.3. Redes Neuronales

Para el caso de redes neuronales se utilizó el método de entrenamiento `trainlm` de Matlab [13] que ajusta pesos y sesgos por defecto, según la función Levenberg-Marquardt, con los valores definidos. Para ver propiedades de la Red, ver anexo A, al final, donde se muestran los comandos en Matlab.

Se trabajó con el 90% para entrenamiento y se reservó el 10% para test.

Este 90% de datos nuevamente se divide de forma aleatoria y se almacena en tres variables que tienen por defecto los siguientes valores (0.70, 0.15 y 0.15). Esta división se obtiene por la función `divideParam` y sus valores en fracción pueden ser mostrados ó modificados mediante los siguientes comandos:

- `net.divideParam.trainRatio`
- `net.divideParam.valRatio`
- `net.divideParam.testRatio`

El parámetro μ inicial es de 0.001 con factor de disminución 0.1 y factor de aumento 10 y el valor máximo de μ es $1e10$.

El rendimiento se mide en MSE. El registro de entrenamiento se puede obtener mediante el comando `tr`. Se realiza en entrenamiento incrementando el número de neuronas en el rango de (1, 20).

Introduciendo el nombre de la red creada y entrenada en la línea de comandos, se puede obtener dimensiones, conexiones, datos (entradas, capas, bias, etc.), funciones y métodos.

Finalmente se prueba el modelo con el 10% de los datos no usados en el entrenamiento. El diagrama de flujo de Redes Neuronales se presenta en la Figura 45.

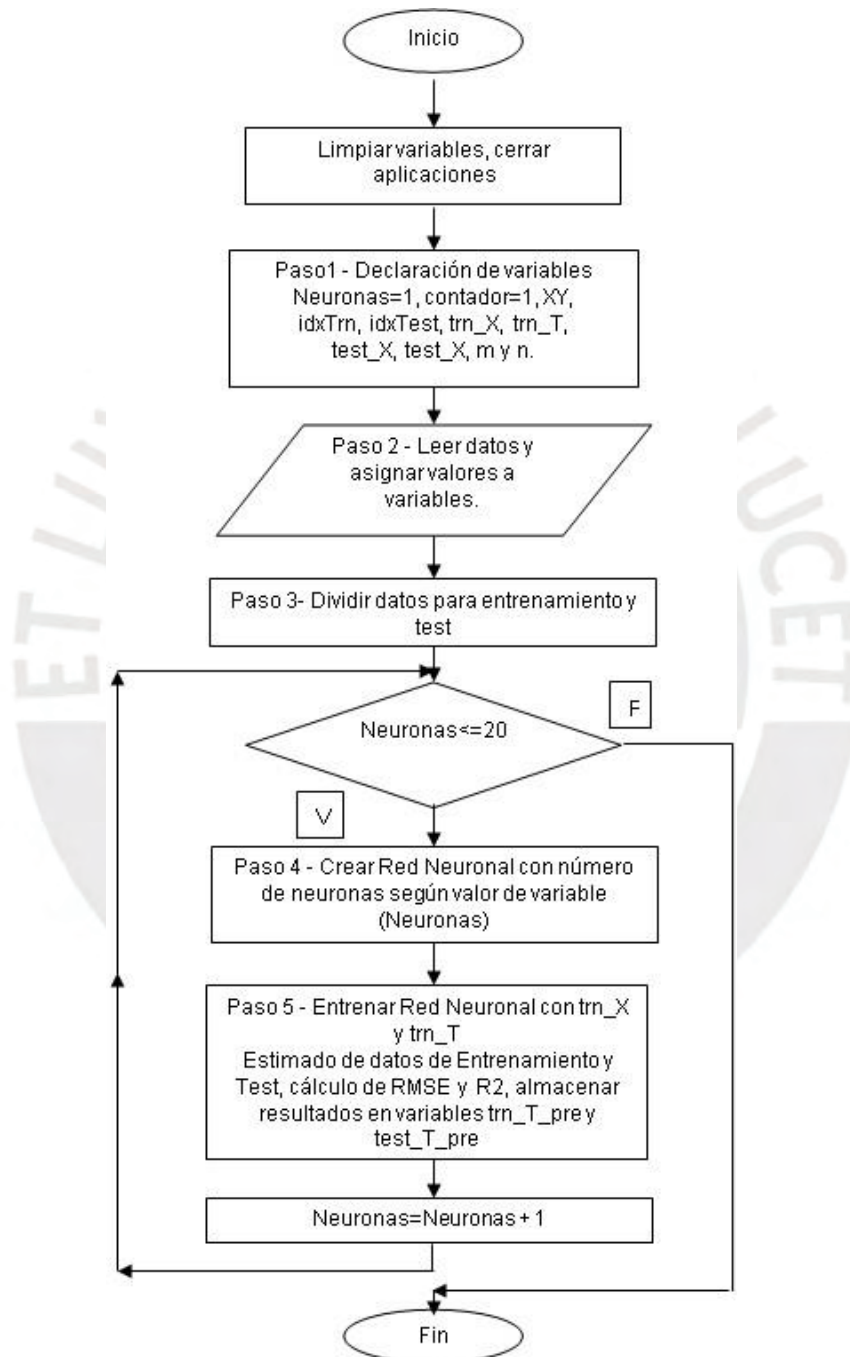


Figura 45 - Diagrama de flujo para algoritmo con Red Neuronal con variación de neuronas en la capa oculta.

El paso 5 de entrenamiento tiene el siguiente diagrama de flujo (Figura 46):

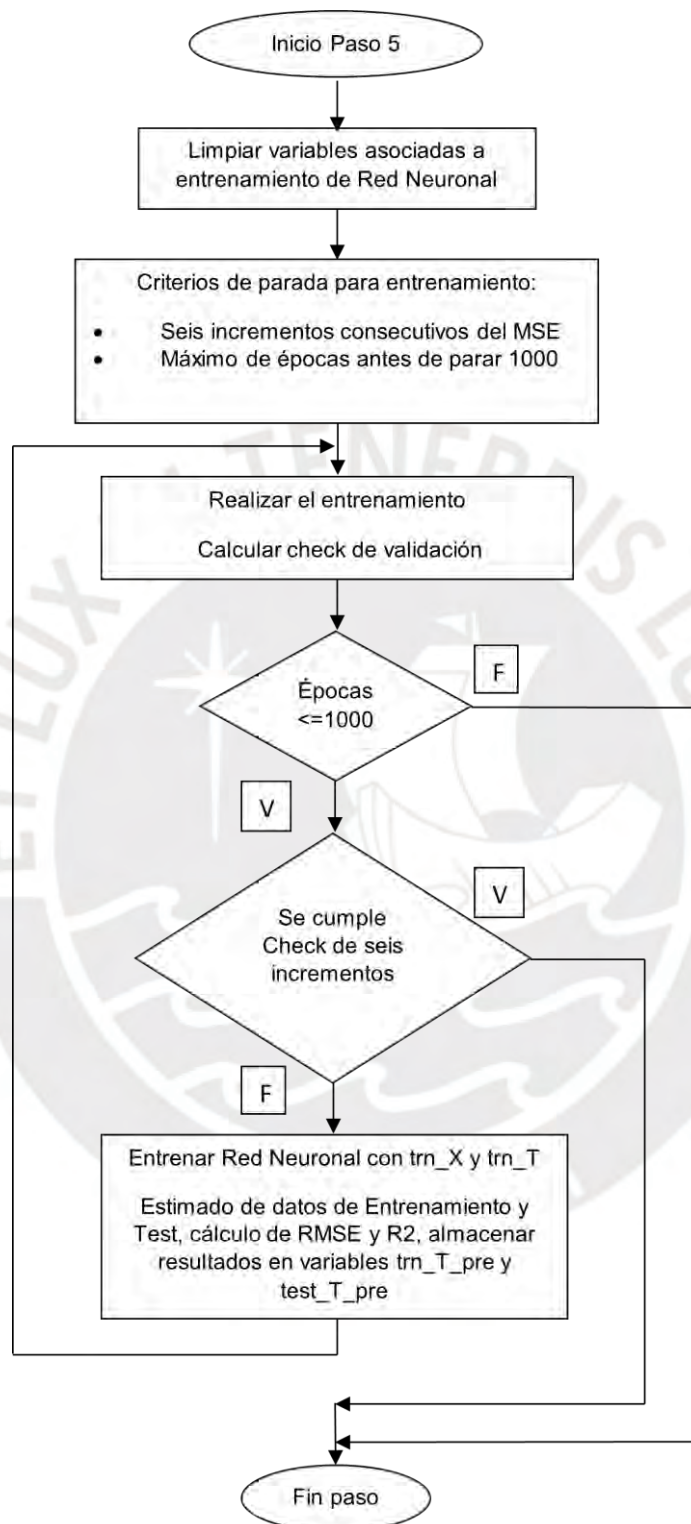


Figura 46 - Diagrama de flujo para criterio de parada de Red Neuronal.

CAPÍTULO 4

RESULTADOS Y ANÁLISIS

Para tener una idea de los tiempos de ejecución de cada modelo desde el inicio hasta su convergencia, se debe tener en cuenta que los modelos se ejecutaron en un equipo con procesador de Intel(R) Core (TM) i5-10400F CPU @ 2.90GHz 2.90 GH y RAM instalada de 16GB. Con esta información mostramos los resultados para cada caso.

4.1. Modelos para ajustar emisiones de CO₂

Las variables predictoras son:

- Presión absoluta en el colector de admisión (kPa);
- Temperatura en la entrada al múltiple de admisión (°C);
- Régimen de giro del motor en RPM;
- Flujo de aire de admisión del motor (g/s);
- Torque efectivo aplicado al motor (Nm).

La variable dependiente es

- Flujo instantáneo de CO₂ (g/s)

4.1.1. Regresión Lineal Múltiple

Se tuvo 21986 registros de 5 variables predictoras y una variable objetivo o dependiente (emisión de CO₂ en g/s) y se aplicaron varios modelos de regresión lineal múltiple para el estimado de emisiones. Los datos para entrenamiento fueron el 90% y para test 10%. Para evaluar los modelos se utilizó RMSE de datos de entrenamiento y de test, tiempo de ejecución y R².

Los resultados mostrados en la Tabla 5, indican que, aun cuando todos los modelos muestran un R² mayor que 0.96, los modelos “Interacciones”, “quadratic” y “poly11222” muestran los mejores resultados con un RMSE casi 8 veces menor en comparación de los modelos “Linear” y “purequadratic”.

Tabla 5 - Resultados para ajuste de emisión de CO₂, modelo de Regresión Lineal Múltiple.

Modelo	RMSE- Entrenamiento	RMSE- Test	tiempo de ejecución (s)	R ² - Entrenamiento	R ² Test
linear	0.235	0.2408	2.42	0.9694	0.9673
interactions	0.0285	0.0285	2.43	0.9995	0.9995
purequadratic	0.2115	0.2154	2.43	0.9752	0.9738
quadratic	0.0284	0.0285	2.58	0.9996	0.9995
poly11222	0.0285	0.0285	2.46	0.9996	0.9995

La Figura 47 muestra la correlación para el modelo “quadratic”. Esta correlación es para los datos de entrenamiento, datos observados y los datos estimados. La línea azul con pendiente de 45° muestra una alta correlación para el caso del modelo. La Figura 48 muestra la correlación para datos de

test y se observa que al igual que para datos de entrenamiento, el coeficiente de correlación confirma una alta correlación.

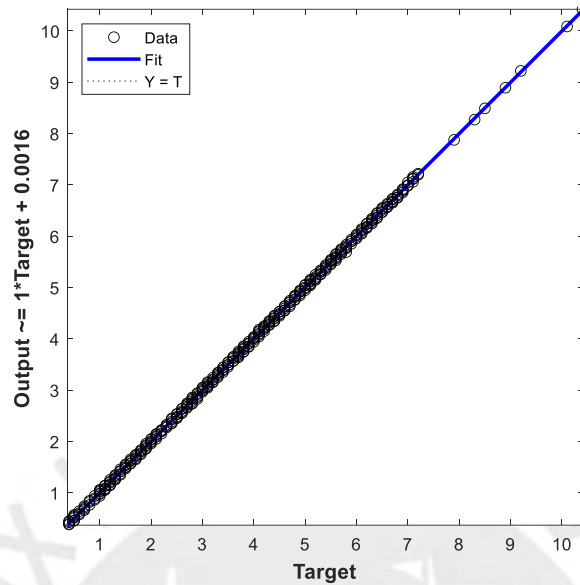


Figura 47 - Datos Observados y estimados para CO₂ con datos de entrenamiento (modelo 'quadratic').

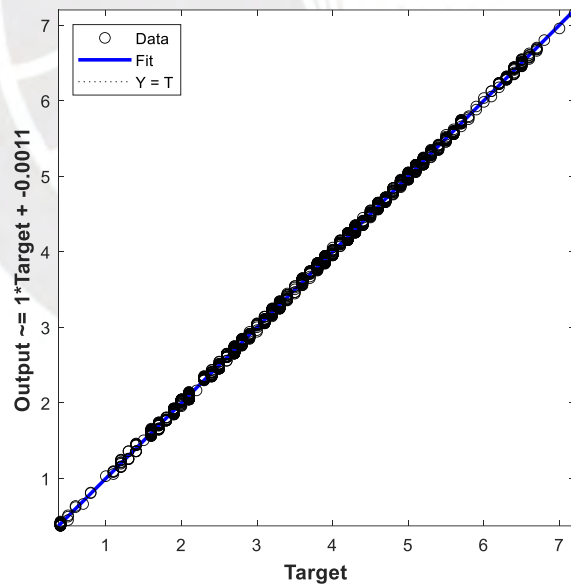


Figura 48 - Datos Observados y estimados para CO₂ con datos de test (modelo 'quadratic').

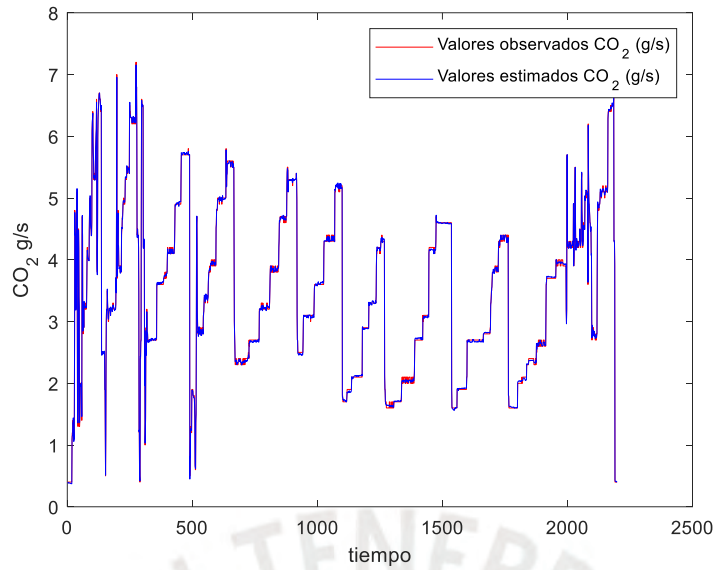


Figura 49 - Comparación valores observados y estimados para emisión de CO₂ para datos de Test.

La Figura 49 muestra de la similitud entre los valores observados y los valores ajustados para las emisiones de CO₂. El error relativo entre el valor real nos da el error relativo, graficamos este error relativo; en un diagrama de cajas (Figura 50).

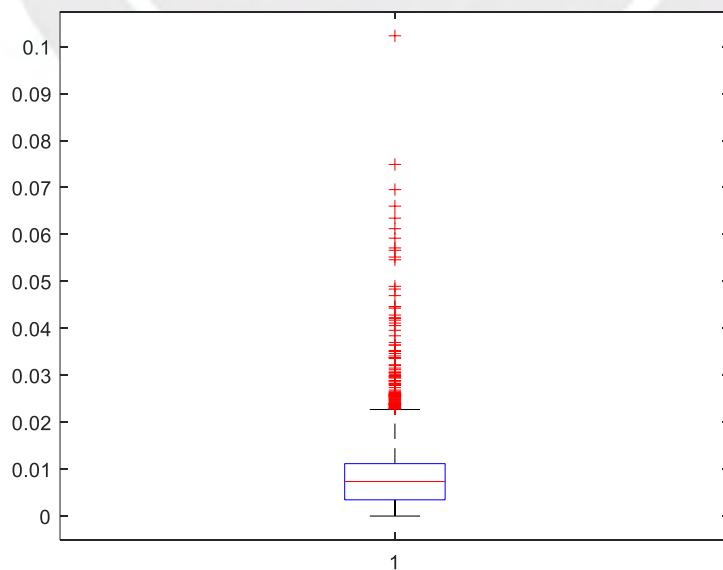


Figura 50 - Diagrama de cajas del error relativo con 115 outliers para ajuste de emisiones de CO₂ para datos de test.

Tabla 6 - Datos de diagrama de cajas para Emisión de CO₂ con 115 outliers-RLM para errores absolutos normalizados - datos de test.

Tipo	Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
quadratic	0.00000	0.00346	0.00736	0.00867	0.01116	0.10231

Según los datos de la Tabla 6, el 75% de los ajustes se pueden obtener con hasta 1.11% de error y la mitad de los datos tienen un error inferior a 0.74%, mientras que el 25% tiene un error de hasta 0.34%.

La Figura 50 muestra los 115 outliers, donde el valor máximo es de 0.1023, es decir que el error absoluto estandarizado máximo es de 10.23%

Utilizando la distancia de Cook y eliminando los registros para los que la distancia de Cook > 3 media del promedio de la distancia de Cook, tenemos como resultado la Figura 51.

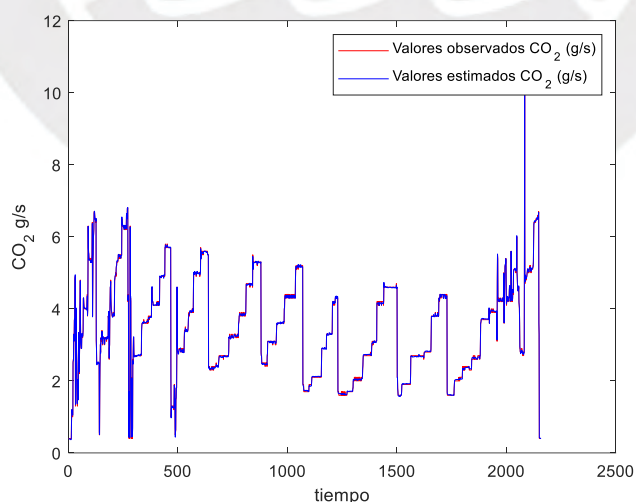


Figura 51 - Comparación valores observados y estimados para emisión de CO₂ sin registros que tienen una distancia de Cook > 3 veces media distancia de Cook.

La Figura 52 muestra los 105 outliers, de los que el valor máximo es de 0.1546. Para esto se eliminó los registros de datos que entregaban la distancia de distancia de Cook >3 media de la distancia de Cook. Aun cuando disminuye el número de Outlier de 115 a 105, crece el valor máximo del error relativo de 0.1023 a 0.1546 (Tabla 7).

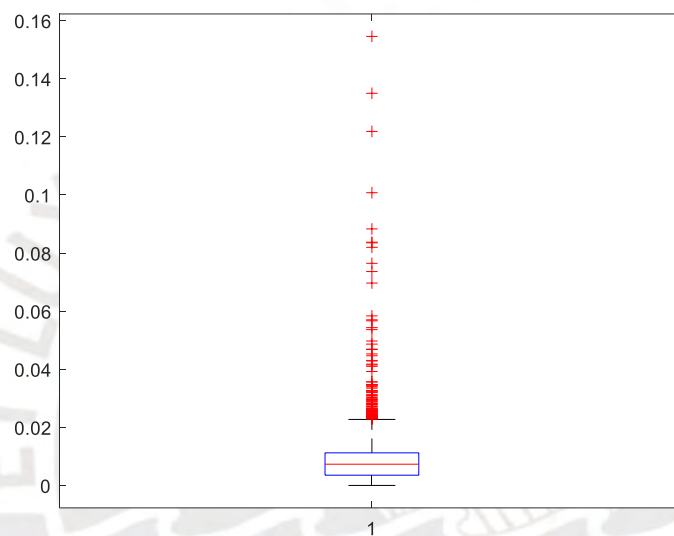


Figura 52 - Diagrama de cajas del error relativo con 105 outliers para ajuste de emisiones de CO_2 para datos de test.

Tabla 7 - Datos de diagrama de cajas para Emisión de CO_2 con 105 outliers-RLM para datos de test.

Tipo	Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
quadratic (3Cook)	0.00001	0.00354	0.00732	0.06739	0.01123	0.15458

Según los datos de la Tabla 7, el 75% de los ajustes se pueden obtener con hasta 1.12% de error y la mitad de los datos tienen un error inferior a 0.73%, mientras que el 25% tiene un error de hasta 0.35%. Si comparamos los datos con la Tabla 6, vemos que el 75% de los datos es similar, pues, de tener el 1.116% de error pasan a tener 1.12%.

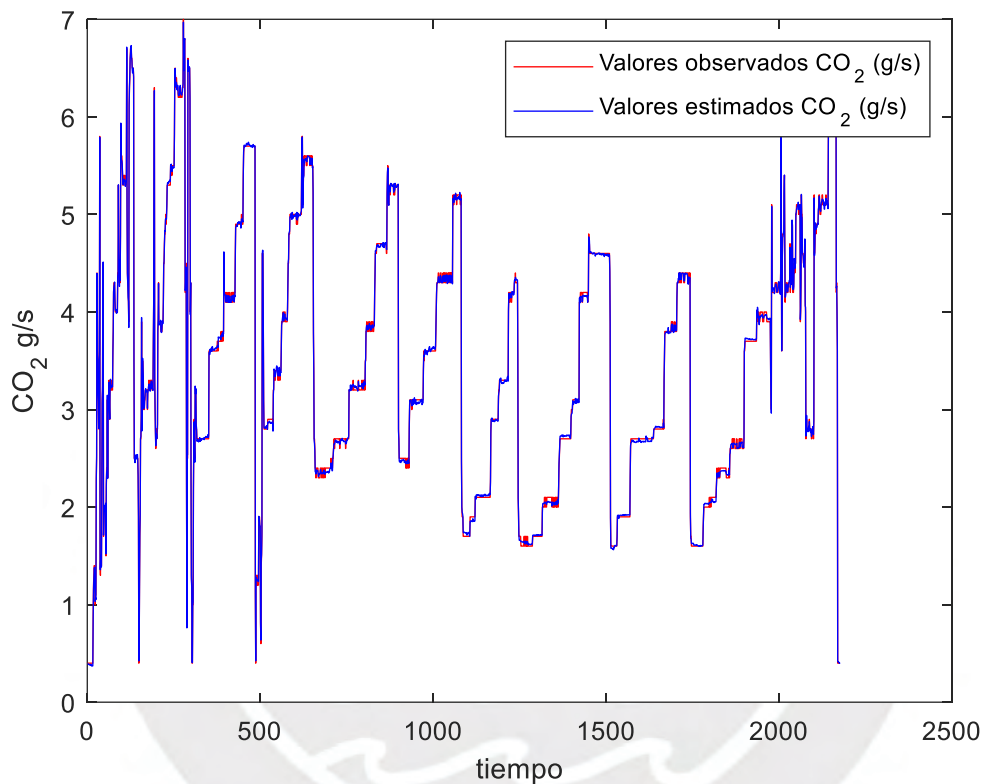


Figura 53 - Comparación valores observados y estimados para emisión de CO₂ con registros que tienen una distancia de Cook < 5 media distancia de Cook.

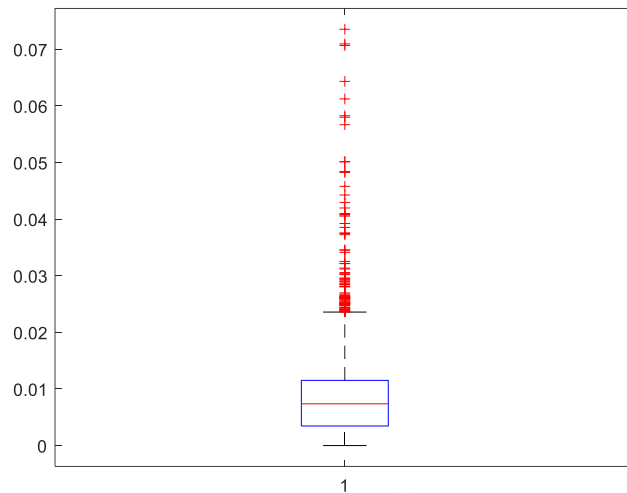


Figura 54 - Diagrama de cajas del error relativo con 97 outliers para ajuste de emisiones de CO_2 -Cook ≤ 5 veces media distancia de Cook para datos de test

Tabla 8-Datos de diagrama de cajas para Emisión de CO_2 con 97 Outlier-MDL

Tipo	Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
quadratic (5Cook)	0.00001	0.00347	0.00738	0.00923	0.01152	0.07355

Se observa que los datos estimados disminuyen su error relativo cuando se retira los registro para datos que tienen distancia de Cook > 5 veces la media de esta distancia. De la Tabla 11 se tiene que el máximo error de todos los datos estimados es de 7.35% y el 75% de los datos tiene un error del 1.15% y el 50% con un error de menos del 0.89%.

La tabla 9 muestra los valores de R^2 similares para los modelos quadratic con datos de diferentes distancias de Cook

Tabla 9 - Resultados con herramienta distancia de Cook.

Modelo	Entrenamiento RMSE	Test RMSE	tiempo de ejecución (s)	R ² Entrenamiento	R ² Test
quadratic	0.0284	0.0285	2.58	0.9996	0.9996
Quadratic (3Cook)	0.0285	0.0285	2.62	0.9995	0.9995
Quadratic (5Cook)	0.0284	0.0286	6.27	0.9995	0.9995

En la Tabla 10 se presentan los resultados del diagrama de cajas para el modelo quadratic. El mejor resultado para el error máximo es el tercero, y éste, considera los datos con distancia de Cook menor o igual a 5 veces el promedio de distancia de Cook

Tabla 10 - Resumen Tablas 6, 7 y 8.

Tipo	Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
quadratic	0.00000	0.00346	0.00736	0.00867	0.01116	0.10231
quadratic (3Cook)	0.00001	0.00354	0.00732	0.06739	0.01123	0.15458
quadratic (5Cook)	0.00001	0.00347	0.00738	0.00923	0.01152	0.07355

4.1.2. Máquina de Soporte Vectorial

Los resultados para SVM en el caso de CO₂, el mejor resultado es para rsvm con parámetros Standarize, kernel polinomial de orden 2 con valor de RMSE de 0.0491 y un R² de 0.999.

Tabla 11 - Resultado para SVM emisiones de CO₂

Máquina de Soporte Vectorial	Entrenamiento RMSE	Test RMSE	Iteraciones	tiempo de ejecución (s)	converge	R ² Entrenamiento	R ² Test
Standardize (false)	2.2802	2.3067	1.00E+06	161.47	no	N/A	N/A
Standardize (true)	0.2396	0.2447	123180	17.51	sí	0.968	0.966
Standardize, Kernel polinomial de orden 2.	0.0491	0.0481	60893	8.41	sí	0.999	0.999
Standardize (true, kernelFunction (gaussian))	0.0973	0.0677	994	1.17	sí	0.995	0.997

El gráfico 55 muestra la correlación para el subconjunto de datos de entrenamiento. Este modelo fue creado estandarizando datos.

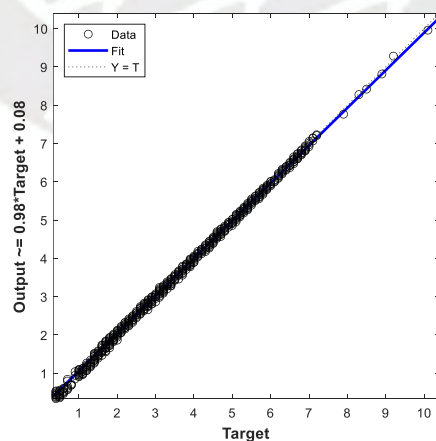


Figura 55 - Comparación valores observados y estimados para CO₂, Standardize, Kernel polinomial de orden 2 (Entrenamiento) SVM.

La predicción de modelo para datos de test como se muestra en la Figura 56, muestra una buena correlación, la pendiente a 45° confirma esto.

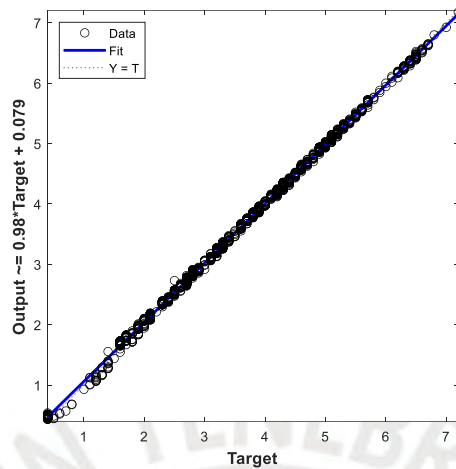


Figura 56 - Comparación valores observados y estimados para CO₂, Standardize, Kernel polinomial de orden 2 (Test) SVM.

La Figura 57 muestra valores estimados comparado con observados para emisiones de CO₂ para datos de test, este modelo usa datos estandarizados.

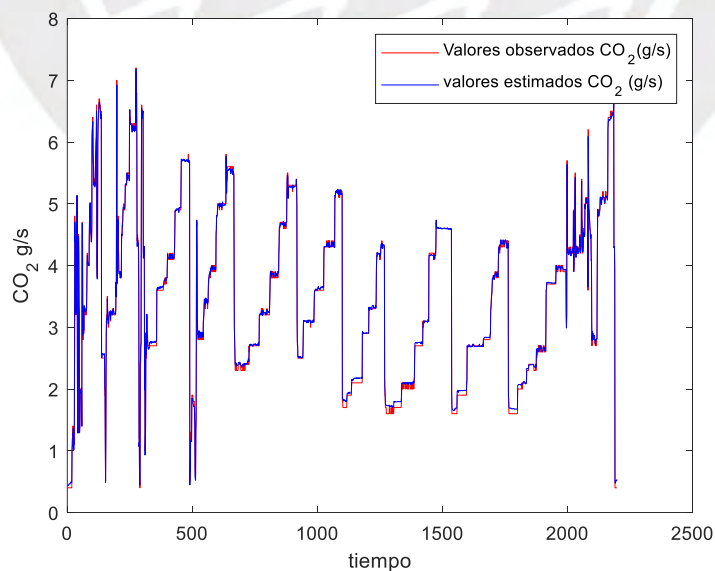


Figura 57 - Valores observados versus estimados para emisiones de CO₂- Standardize, Kernel polinomial de orden 2 SVM.

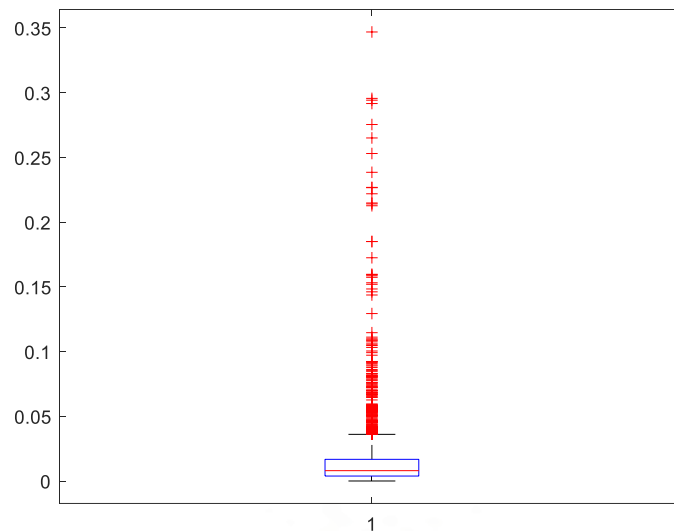


Figura 58 - Diagrama de cajas para error relativo- Standarize, Kernel polinomial de orden 2 con 293 outliers para datos de test.

Tabla 12 - Datos de diagrama de cajas para Emisión de CO₂ -SVM- Standarize, Kernel polinomial de orden 2.

Tipo	Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
Standardize, Kernel polinomial de orden 2.	0.00001	0.00383	0.00799	0.01683	0.01673	0.34692

De los datos de la Tabla 12 que corresponden a la Figura 58 se puede ver que para el ajuste para emisiones de CO₂ usando SVM, en el 75% de datos tiene un error de 1.67%, mientras que para el 50% de los datos tiene un error relativo del 0.799% y finalmente para el 25% de los datos el error sería de hasta 0.382%.

4.1.3. Redes neuronales

Se ejecuta el modelo con la función de activación tansig y algoritmo de entrenamiento Levenberg-Marquardt, variando la cantidad de neuronas de la capa oculta en el rango de 1 al 20. Con los resultados se realiza una comparación entre los RMSE y el número de neuronas.

Tabla 13 - Valores de RMSE para arreglos de 1 a 20 neuronas.

Redes Neuronales	RMSE Entrenamiento	RMSE Test	R ² Entrenamiento	R ² Test
1 neurona	0.1472	0.1543	0.9880	0.9866
2 neuronas	0.0365	0.0411	0.9993	0.9990
3 neuronas	0.0329	0.0337	0.9994	0.9994
4 neuronas	0.0283	0.0284	0.9996	0.9995
5 neuronas	0.0347	0.0419	0.9993	0.9990
6 neuronas	0.0278	0.0277	0.9996	0.9996
7 neuronas	0.0280	0.0279	0.9996	0.9996
8 neuronas	0.0281	0.0283	0.9996	0.9995
9 neuronas	0.0276	0.0276	0.9996	0.9996
10 neuronas	0.0277	0.0276	0.9996	0.9996
11 neuronas	0.0277	0.0277	0.9996	0.9996
12 neuronas	0.0277	0.0278	0.9996	0.9996
13 neuronas	0.0265	0.0267	0.9996	0.9996
14 neuronas	0.0272	0.0272	0.9996	0.9996
15 neuronas	0.0276	0.0275	0.9996	0.9996
16 neuronas	0.0274	0.0278	0.9996	0.9996
17 neuronas	0.0270	0.0272	0.9996	0.9996
18 neuronas	0.0267	0.0269	0.9996	0.9996
19 neuronas	0.0272	0.0271	0.9996	0.9996
20 neuronas	0.0277	0.0275	0.9996	0.9996

La Figura 59 muestra un menor RMSE (entrenamiento) para la capa con 13 neuronas en la predicción de CO₂. También aprecia que la precisión no mejora al incrementar el número de neuronas de la capa oculta. Se puede ver que no hay una relación directa (a más neuronas no asegura más precisión) entre la disminución del RMSE y el número de neuronas.

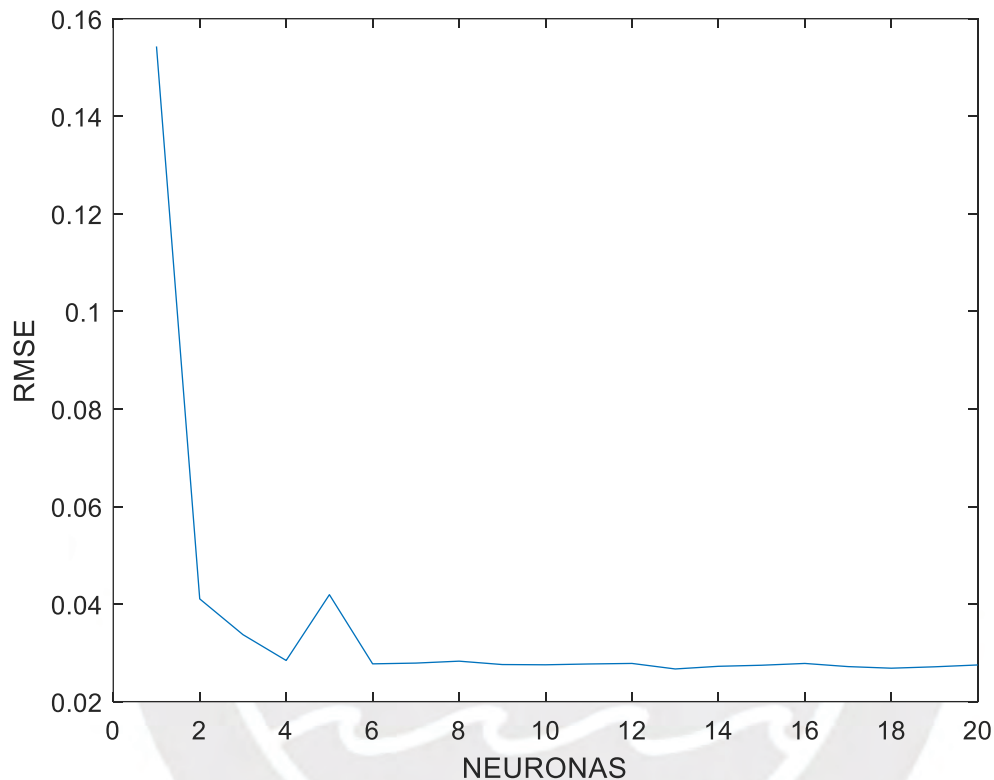


Figura 59 - RMSE (entrenamiento) del flujo de CO₂ en función de cantidad de neuronas para estimado de emisión de CO₂

Podemos analizar los resultados con el diagrama de bigotes, ver Tablas 14, 15 y 16

Tabla 14 - Datos de caja de bigotes para errores absolutos normalizados 1 a 7 neuronas-test-Estimado CO₂.

Datos de Caja de Bigotes	1 neurona	2 neuronas	3 neuronas	4 neuronas	5 neuronas	6 neuronas	7 neuronas
Mínimo	0.000%	0.000%	0.001%	0.000%	0.000%	0.001%	0.000%
Primer Cuartil	1.271%	0.398%	0.328%	0.323%	0.382%	0.303%	0.308%
Mediana	2.306%	0.821%	0.720%	0.694%	0.782%	0.651%	0.666%
Media	3.452%	1.202%	1.015%	0.838%	1.162%	0.825%	0.821%
Tercer cuartil	3.979%	1.330%	1.237%	1.098%	1.254%	1.061%	1.083%
Máximo	127.381%	39.408%	14.981%	11.226%	47.319%	10.934%	12.041%

Tabla 15 - Datos de caja de bigotes para errores absolutos normalizados 8 a 14 neuronas-test-Estimado de CO₂.

Datos de Caja de Bigotes	8 neuronas	9 neuronas	10 neuronas	11 neuronas	12 neuronas	13 neuronas	14 neuronas
Mínimo	0.001%	0.001%	0.001%	0.001%	0.001%	0.001%	0.000%
Primer Cuartil	0.309%	0.281%	0.293%	0.296%	0.301%	0.262%	0.261%
Mediana	0.673%	0.647%	0.652%	0.648%	0.668%	0.583%	0.601%
Media	0.833%	0.795%	0.801%	0.803%	0.826%	0.758%	0.775%
Tercer cuartil	1.083%	1.052%	1.041%	1.050%	1.044%	1.000%	1.018%
Máximo	12.689%	9.640%	11.806%	9.425%	10.766%	7.848%	8.229%

Tabla 16 - Datos de caja de bigotes para errores absolutos normalizados 15 a 20 neuronas-test-Estimado de CO₂.

Datos de Caja de Bigotes	15 neuronas	16 neuronas	17 neuronas	18 neuronas	19 neuronas	20 neuronas
Mínimo	0.000%	0.000%	0.001%	0.001%	0.000%	0.000%
Primer Cuartil	0.299%	0.280%	0.264%	0.268%	0.270%	0.286%
Mediana	0.648%	0.641%	0.603%	0.592%	0.618%	0.653%
Media	0.794%	0.795%	0.767%	0.745%	0.767%	0.790%
Tercer cuartil	1.030%	1.035%	1.007%	0.997%	1.021%	1.034%
Máximo	7.983%	8.285%	9.367%	8.930%	6.125%	7.332%

Graficando los datos de las Tablas (15, 16 y 17) podemos tener una idea más clara sobre los datos de la caja de bigotes. La red con 19 neuronas tiene un error máximo de 6.125% y la red con 13 neuronas tiene 7.848%

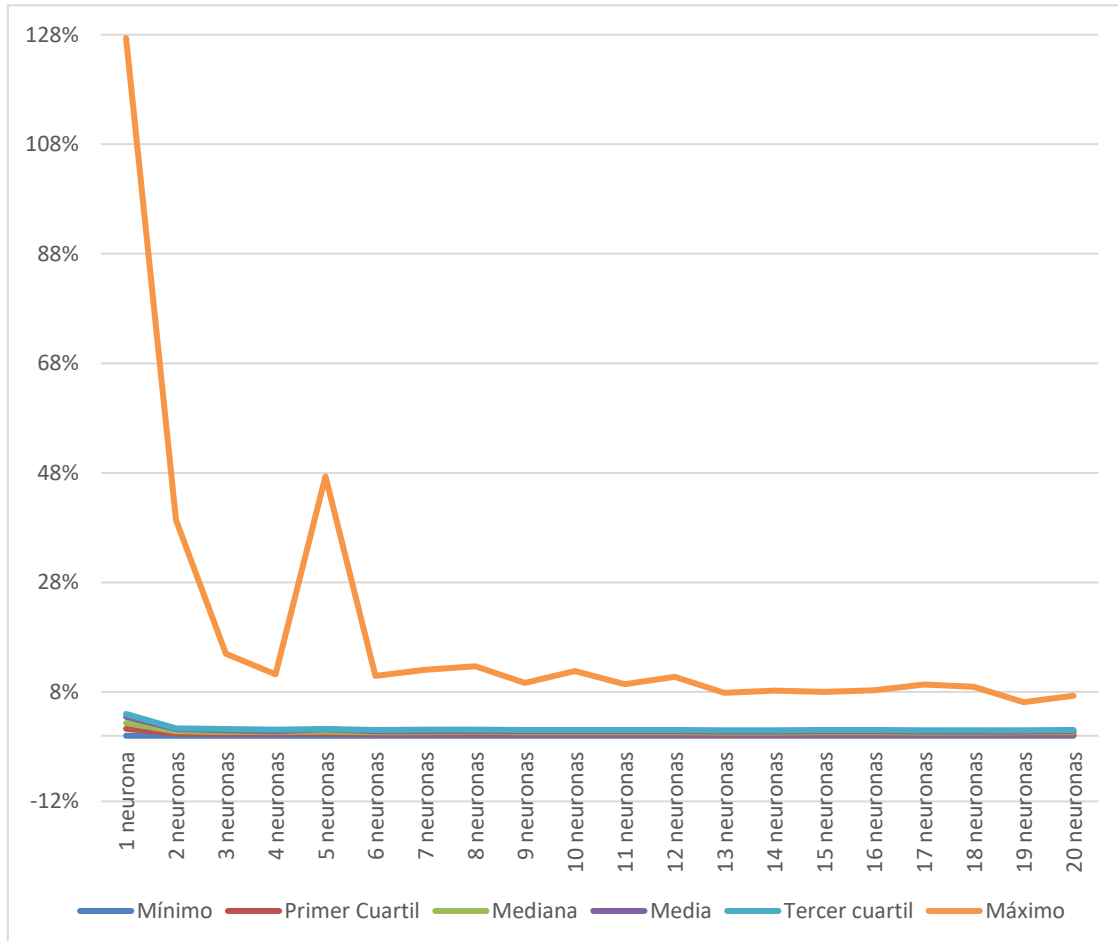


Figura 60 - Evolución de error relativo estimado de CO₂ datos de test - Tablas 18,19 y 20

El error relativo para estimado de emisiones de CO₂ llega a su punto más bajo para 19 neuronas (Figura 60). Sin embargo, el menor RMSE de entrenamiento y test se da con 13 neuronas (Tabla 13).

La Figura 60 muestra la evolución de RMSE de entrenamiento y de Test para arreglos de dos capas y 1 a 20 neuronas en la capa oculta.

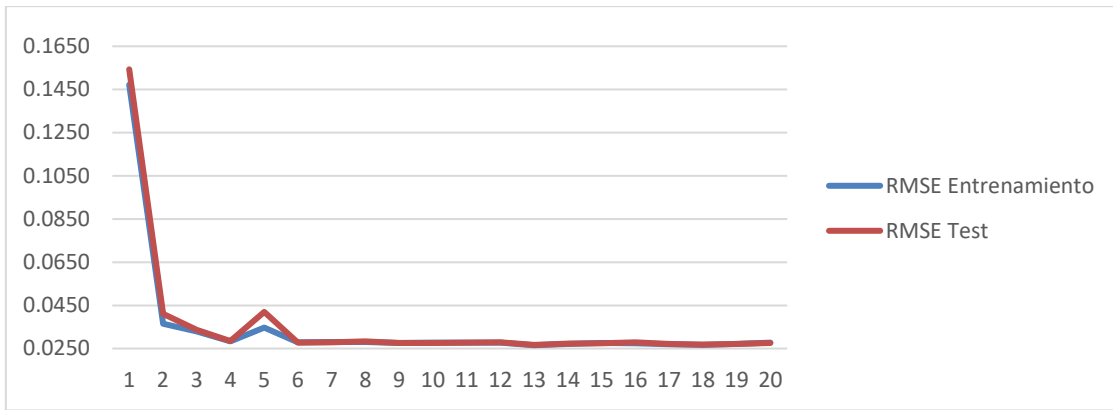


Figura 61 - RMSE para 1 a 20 neuronas para estimado de emisiones de CO₂

El MSE para test es similar al de validación y entrenamiento. La mejor validación se muestra en la época 191, como muestra la Figura 62.

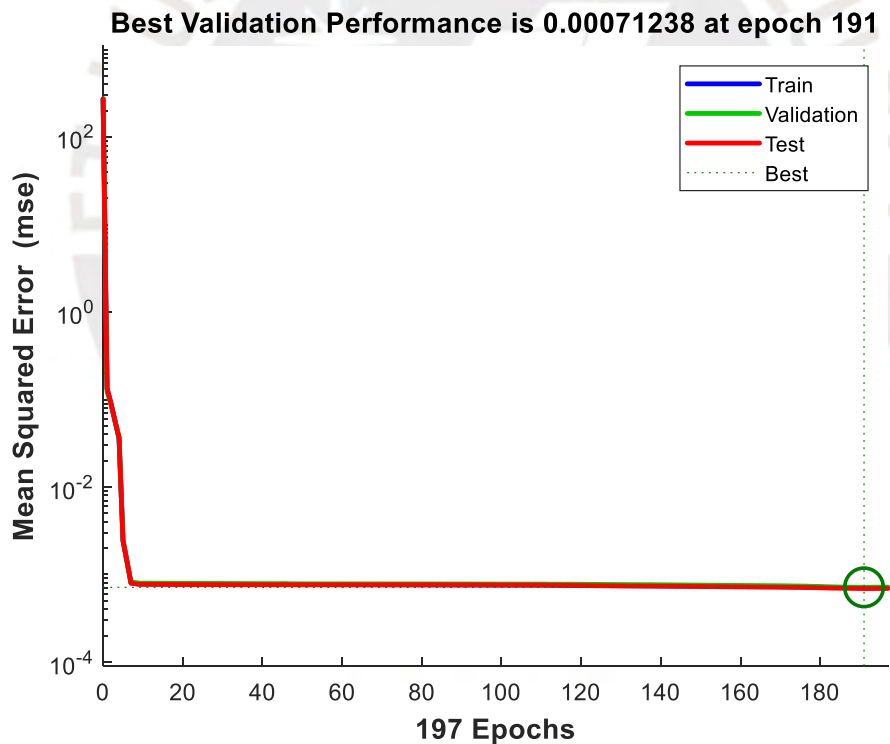


Figura 62 - MSE para entrenamiento, validación y test con 13 neuronas para estimado de emisión de CO₂

La época 197 muestra que se cumple el check de validación y el programa realiza la parada, dando el mejor MSE para la época 191, ver Figura 63.

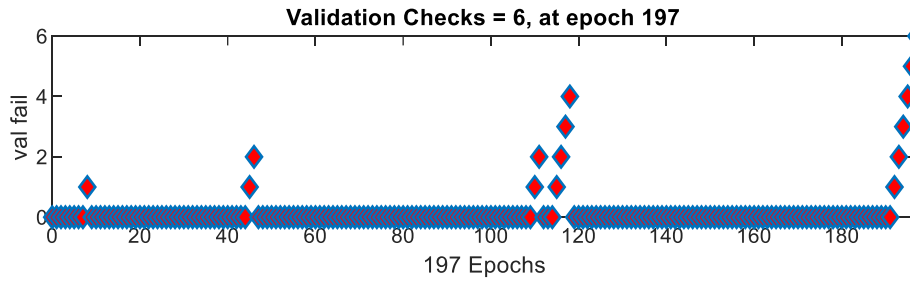


Figura 63 - Entrenamiento de la red con 13 neuronas para predicción de CO₂ con Redes Neuronales

La Figura 64 muestra los diagramas de correlación para los diferentes datos, entrenamiento con $R^2=0.9996$ y test con $R^2=0.9996$ (ver Tabla 13) para red con 13 neuronas en la capa oculta. Todos muestran muy buen ajuste con un R^2 cercano a 1 y el pendiente de 45° entre datos observados y estimados

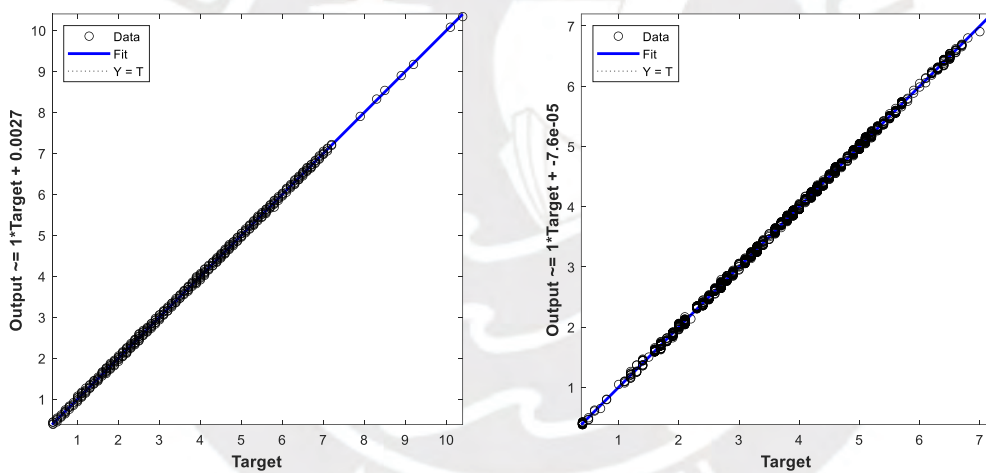


Figura 64 - Diagramas de correlación para predicción de emisiones de CO₂ Redes Neuronales, para datos de entrenamiento (izquierda) y test (derecha).

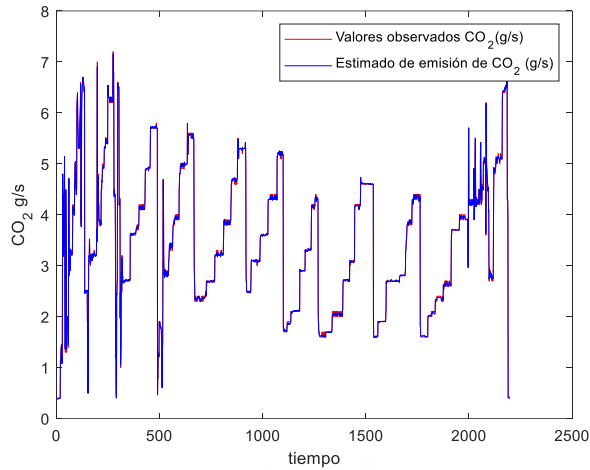


Figura 65 - Valores observados versus predicción para emisiones de CO₂ con 13 neuronas en la capa oculta

De los datos de la Tabla 17 que corresponden al diagrama de caja de la Figura 66, se extrae que, el 75% de los datos tiene como máximo un error de 1.00%, mientras que el 50% tiene 0.583% y el 25% tiene como máximo un error de 0.26% y error relativo máximo de 7.85%.

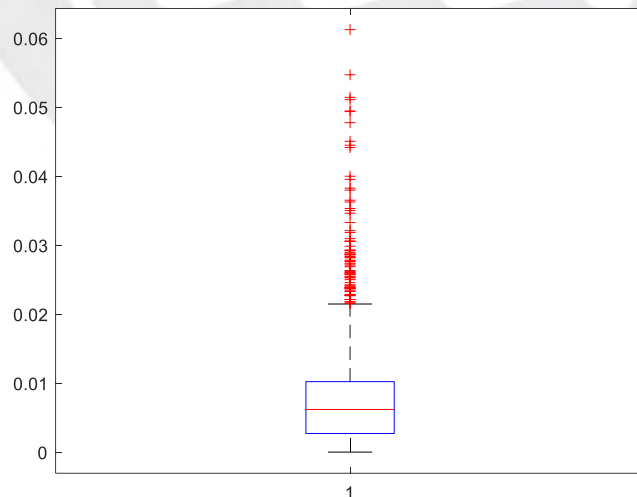


Figura 66 - Diagrama de cajas de error relativo con 100 outliers, red con 13 neuronas en la capa oculta.

Tabla 17 - Datos de diagrama de cajas para Emisión de CO₂ usando Redes Neuronales.

Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
0.0000	0.0026	0.0058	0.0076	0.0100	0.0785

Podemos usar la herramienta usada en Regresión Lineal Múltiple para ver el impacto en el modelo de Redes Neuronales. Usando los registros filtrados por distancia de Cook > 5 veces el promedio de esta distancia tenemos el siguiente resultado con una mejora en el número de outliers de 100 a 95 y una disminución en el error absoluto normalizado máximo de 7.85% a 7.09%

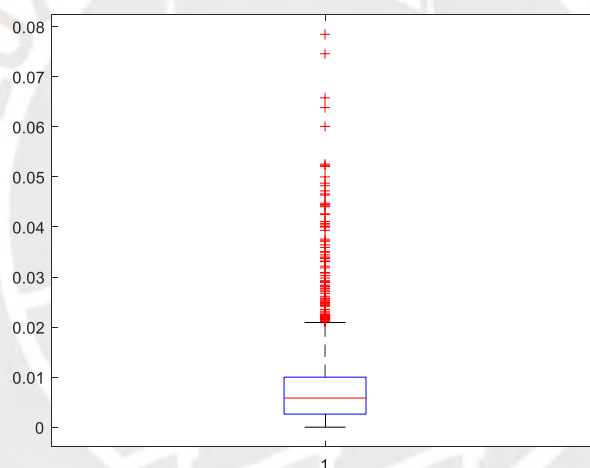


Figura 67 - Diagrama de cajas de error relativo para Test, con 95 outliers para emisiones de CO₂ usando datos filtrados por distancia de Cook, Red Neuronal con 13 neuronas en la capa oculta.

Tabla 18 - Datos de diagrama de cajas para emisiones de CO₂ usando datos filtrados de distancia de Cook

Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
0.000%	0.310%	0.662%	0.827%	1.127%	7.085%

El máximo error para este caso es 7.085% (usando distancia de Cook).

4.2. Modelos para ajustar el consumo de combustible

4.2.1. Regresión Lineal Múltiple

La Tabla 19 muestra los resultados para la estimación del consumo de combustible y los modelos que mejor se ajustan son: “interactions”, “quadratic” y “poly11222”, con un R^2 de 0.9998 durante el entrenamiento.

Tabla 19 - Resultados para Regresión Lineal Múltiple.

Regresión Lineal Múltiple	Entrenamiento RMSE	Test RMSE	tiempo de ejecución (s)	R^2 Entrenamiento	R^2 Test
linear	0.2789	0.2864	2.44	0.9695	0.9553
interactions	0.0230	0.0228	2.62	0.9998	0.9994
purequadratic	0.2499	0.2549	2.66	0.9755	0.9633
quadratic	0.0230	0.0228	2.59	0.9998	0.9994
poly11222	0.0230	0.0228	2.62	0.9998	0.9994

Analizando los datos de test mediante el diagrama de caja de bigotes (Tabla 20), se puede verificar que “interactions” tiene un menor error absoluto normalizado para test, y está muy cerca de los valores observados.

La Tabla 20 muestra que “interaction” tiene un error absoluto normalizado máximo de 6.692% y el 75% de los datos tienen como máximo un error de 0.77%.

Tabla 20 - Datos de caja de bigotes de errores relativos de datos de test.

Datos de Caja de Bigotes	Lineal	interaction	purequadratic	quadratic	poly11222
Mínimo	0.001%	0.000%	0.004%	0.000%	0.000%
Primer Cuartil	0.872%	0.226%	1.248%	0.226%	0.226%
Mediana	2.205%	0.459%	2.190%	0.454%	0.454%
Media	7.375%	0.584%	6.622%	0.586%	0.586%
Tercer cuartil	5.333%	0.773%	4.401%	0.775%	0.777%
Máximo	258.888%	6.692%	223.632%	6.935%	6.955%

La Figura 68 muestra la correlación para datos de entrenamiento del consumo de combustible observado y el estimado utilizando el modelo 'interaction', Se constata la alta correlación, con $R^2 = 0.999$ (ver Tabla 19)

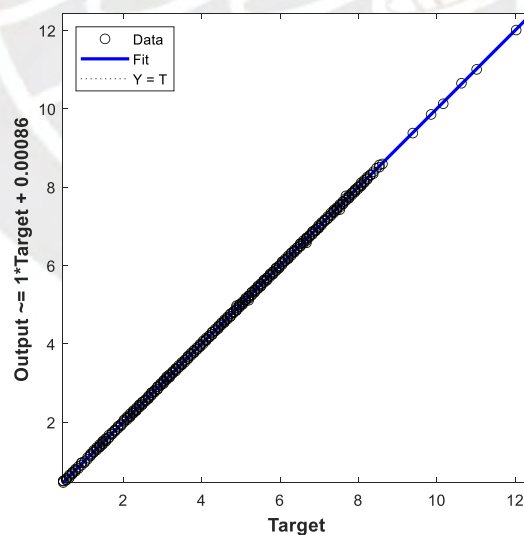


Figura 68- Datos Observados y pronóstico para consumo de combustible, datos de entrenamiento ("interaction").

La Figura 69 muestra una buena correlación ($R^2=0.999$ datos de test) para el modelo “interactions” y es muy similar a la obtenida en los datos entrenamiento. Líneas abajo, la Figura 70 nos muestra un gráfico de los correspondientes datos observados y estimados.

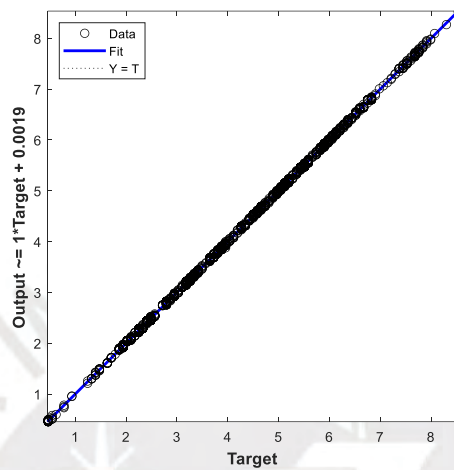


Figura 69 - Datos Observados y pronóstico para consumo de combustible, datos de test (modelo “interactions”).

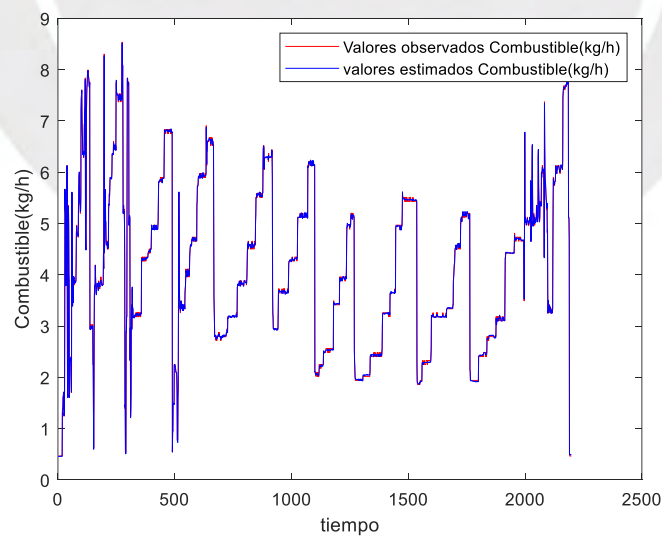


Figura 70 - Comparación valores observados y estimado para consumo de combustible.

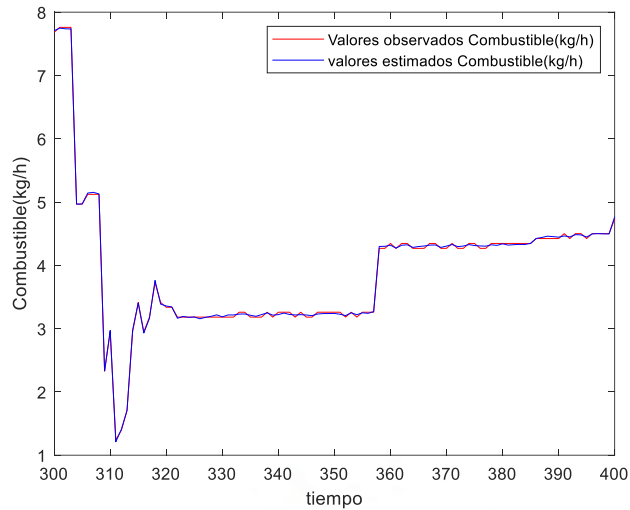


Figura 71 - Valores observados y estimados para el consumo de combustible en el rango de [300,400].

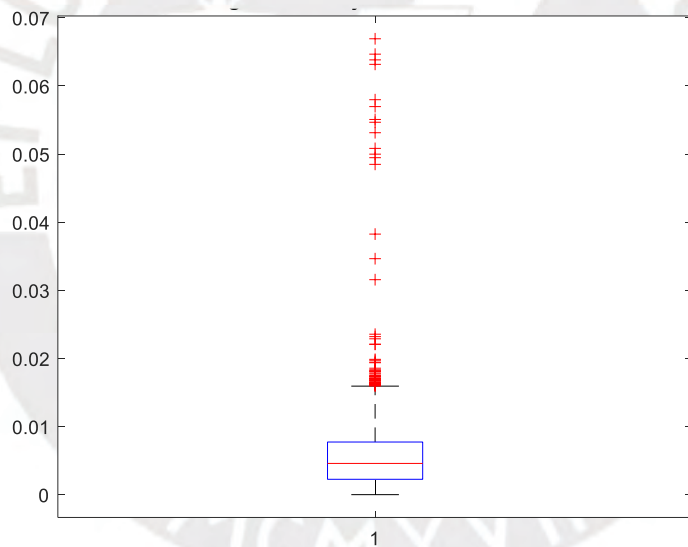


Figura 72 - Diagrama de cajas del error relativo con 211 outliers para ajuste de consumo combustible

Los datos de la Tabla 21 que corresponden al diagrama de cajas de la Figura 72 muestran que; el 75% de los datos poseen un error de hasta 0.773 %; el 50% muestra un error de hasta 0.584%; mientras que el 25% muestra un error

de hasta 0.226%. Una observación a resaltar es que el error relativo máximo es de 6.692%.

Tabla 21 - Datos de diagrama de cajas para consumo de combustible para el modelo de Regresión Lineal Múltiple.

Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
0.00000	0.00226	0.00459	0.00584	0.00773	0.06692



4.2.2. Máquina de Soporte Vectorial

Se realizó la estimación del consumo de combustible en las siguientes condiciones:

- Con datos sin estandarizar;
- Con datos estandarizados;
- Con datos estandarizados, función Kernel polinomial de orden 2;
- Con datos estandarizados y funciona Kernel Gaussiana.

Los resultados se muestran en la Tabla 22. La primera estimación sin estandarizar los datos, no se consiguió la convergencia. De aquí, en adelante, todos los procesos se realizaron estandarizando los datos y con esto puede verse que el menor RMSE se obtiene para Kernel polinomial de orden 2 con un valor de 0.0611 (entrenamiento) y un RMSE de 0.0616 (test), y un R^2 de 0.999

Tabla 22 - Resultado para SVM del consumo de combustible.

Máquina de Soporte Vectorial	Entrenamiento RMSE	Test RMSE	Iteraciones	tiempo de ejecución (s)	converge	R^2 Entrenamiento	R^2 Test
Standarize (false)	1.0229	1.0299	1000000	140.24	NO	N/A	N/A
Standarize (true)	0.2831	0.2902	112434	16.09	SI	0.967	0.966
Standardize, Kernel polinomial de orden 2.	0.0611	0.0616	71402	9.34	SI	0.999	0.999
Standardize (true, kernelFunction (gaussian))	0.1151	0.0797	994	1.03	SI	0.995	0.998

Para confirmar que la mejor estimación es de Kernel polinomial de orden 2, se analiza el resultado de estimación del consumo para los datos de test. Como se puede apreciar en la Tabla 23, éste presenta un error absoluto estandarizado máximo de 35.36%.

Tabla 23 - Resultado para SVM consumo de combustible-Diagrama de cajas.

Datos de Caja de Bigotes	Standarize (false)	Standarize (true)	Standardize, Kernel polinomial de orden 2.	Standardize (true, kernelFunction (gaussian))
Mínimo	0.033%	0.001%	0.001%	0.006%
Primer Cuartil	10.279%	1.123%	0.392%	0.567%
Mediana	19.593%	2.523%	0.931%	1.339%
Media	19.645%	7.895%	1.854%	2.193%
Tercer cuartil	26.029%	5.040%	2.060%	2.636%
Máximo	138.100%	278.396%	35.363%	77.115%

La Figura 73 muestra buena correlación entre datos observados y estimados durante el entrenamiento para el caso de datos estandarizados con función kernel polinomial de orden 2. El valor de R^2 logrado es 0.999 (ver Tabla 22)

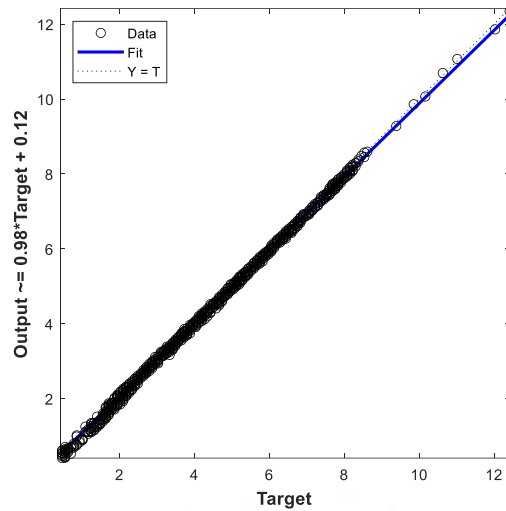


Figura 73 - Comparación valores observados y ajustado para el consumo de combustible, Kernel polinomial de orden2 (Entrenamiento) para SVM.

La Figura 74 muestra la correlación entre datos observados y estimados durante el test. Una pendiente de 45° indica que la correlación entre datos observados y estimados por el modelo Máquina de Soporte Vectorial es un modelo aceptable.

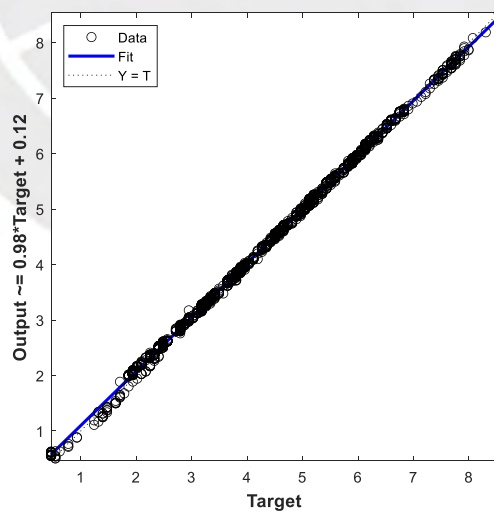


Figura 74 - Comparación valores observados y ajustado para combustible, Kernel polinomial de orden 2 (test) para SVM.

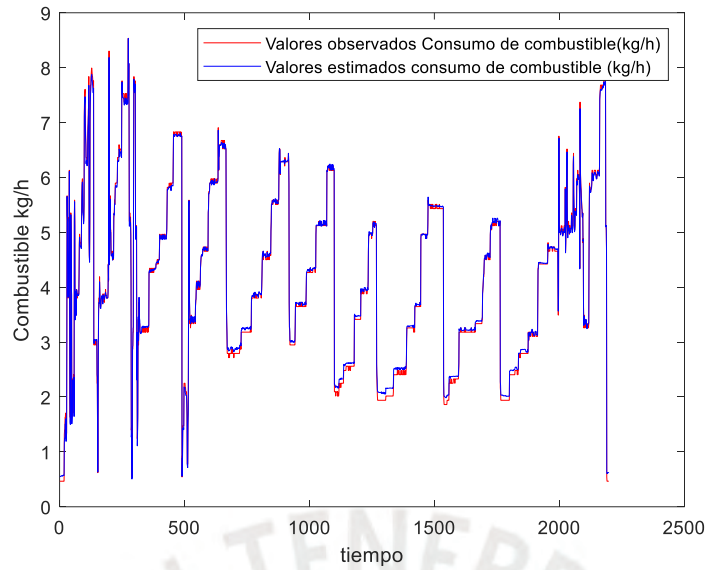


Figura 75 - Comparación valores observados y ajustado para el consumo de combustible con SVM

El diagrama de cajas de muestra 218 outliers (ver Figura 76)

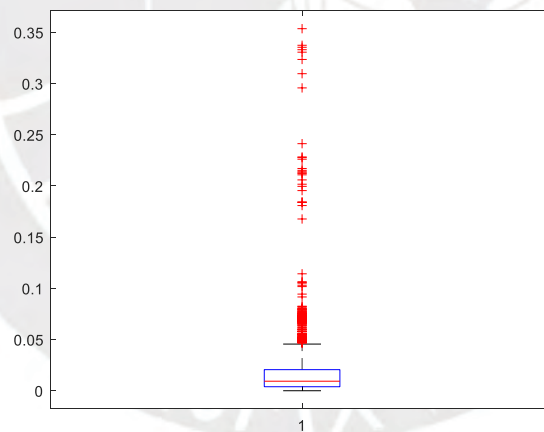


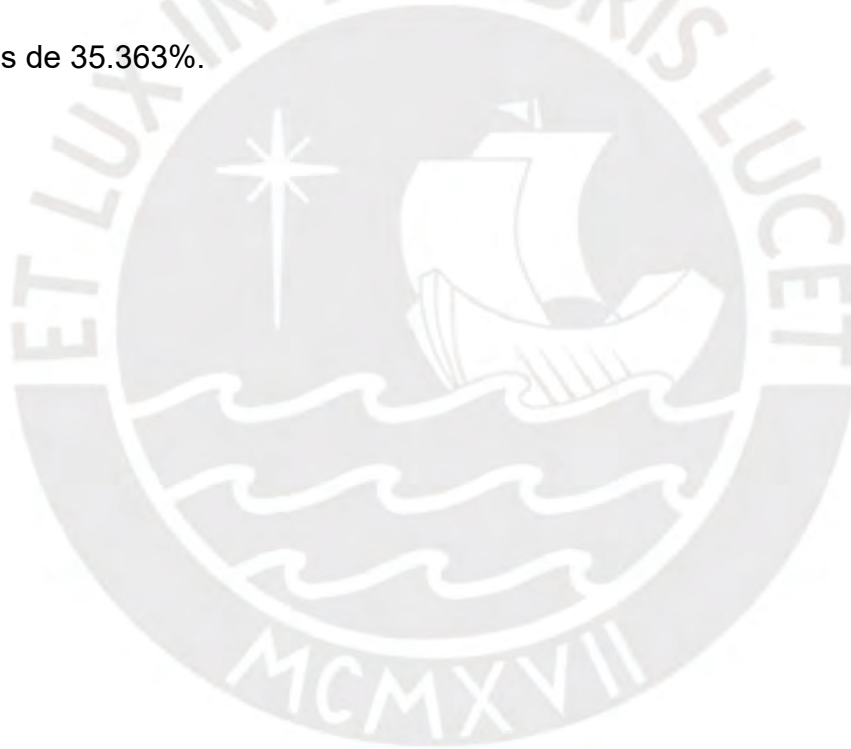
Figura 76 - Diagrama de cajas del error relativo con 218 outliers para ajuste de combustible-SVM

Un total de 218 Outliers de un total de 2198 datos de test representa el 9.92% del total de datos de test, y por lo tanto se requiere encontrar un modelo que minimice esta cantidad.

Tabla 24 - Datos de diagrama de cajas para combustible - SVM

Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
0.000001	0.0039	0.0093	0.01854	0.02060	0.35363

Los datos de la Tabla 24 muestra que para SVM en el caso de consumo de combustible se tiene que el 75% de los datos tiene un error absoluto estandarizado de 2.06%, el 50% tiene valores de error relativo de hasta 2.52% y el 25% tiene errores relativos de hasta 0.39%. El error relativo máximo es de 35.363%.



4.2.3. Redes neuronales

En este modelo se considera varias estimaciones variando el número de neuronas de la capa oculta desde uno hasta veinte. La variación de los RMSE para los diferentes arreglos se muestra en la tabla 25 y el menor valor se consigue con un arreglo de 19 neuronas en la capa oculta

Tabla 25-Variación de RMSE de entrenamiento y test para Redes neuronales en función de la cantidad de neuronas.

Redes Neuronales	RMSE Entrenamiento	RMSE Test	R ² Entrenamiento	R ² Test
1 neurona	0.1733	0.1823	0.9882	0.9867
2 neuronas	0.0343	0.0408	0.9995	0.9993
3 neuronas	0.0258	0.0265	0.9997	0.9997
4 neuronas	0.0230	0.0229	0.9998	0.9998
5 neuronas	0.0301	0.0408	0.9996	0.9993
6 neuronas	0.0229	0.0228	0.9998	0.9998
7 neuronas	0.0230	0.0228	0.9998	0.9998
8 neuronas	0.0230	0.0228	0.9998	0.9998
9 neuronas	0.0229	0.0228	0.9998	0.9998
10 neuronas	0.0241	0.0231	0.9998	0.9998
11 neuronas	0.0227	0.0226	0.9998	0.9998
12 neuronas	0.0228	0.0227	0.9998	0.9998
13 neuronas	0.0221	0.0221	0.9998	0.9998
14 neuronas	0.0227	0.0226	0.9998	0.9998
15 neuronas	0.0229	0.0229	0.9998	0.9998
16 neuronas	0.0225	0.0225	0.9998	0.9998
17 neuronas	0.0229	0.0228	0.9998	0.9998
18 neuronas	0.0221	0.0222	0.9998	0.9998
19 neuronas	0.0219	0.0219	0.9998	0.9998
20 neuronas	0.0236	0.0227	0.9998	0.9998

La Figura 77 muestra un menor RMSE del consumo de combustible y se obtuvo con una red con 19 neuronas en la capa oculta. También notamos que la precisión no mejora al incrementar el número de neuronas de la capa oculta. No hay una relación directa (más neuronas no asegura un resultado con mejor precisión).

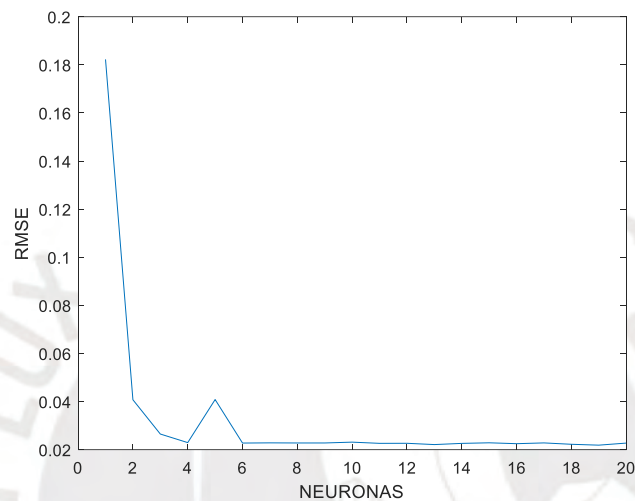


Figura 77 – RMSE (entrenamiento) en función de cantidad de neuronas para estimado de consumo de Gasolina.

A continuación, mostramos el error absoluto normalizado para los veinte escenarios. La Tabla 26 expone los resultados en porcentajes para datos de test de caja de bigotes. De todos ellos el peor escenario corresponde a la red de una neurona y el mejor es la red con 4 neuronas con un error absoluto estandarizado máximo de 6.566%.

Tabla 26 - Datos de caja de bigotes para errores absolutos normalizados 1 a 7 neuronas-test.

Datos de Caja de Bigotes	1 neurona	2 neuronas	3 neuronas	4 neuronas	5 neuronas	6 neuronas	7 neuronas
Mínimo	0.002%	0.001%	0.001%	0.000%	0.000%	0.000%	0.000%
Primer Cuartil	1.203%	0.247%	0.234%	0.222%	0.219%	0.222%	0.220%
Mediana	2.342%	0.524%	0.464%	0.465%	0.485%	0.450%	0.462%
Media	3.426%	0.912%	0.663%	0.592%	0.811%	0.589%	0.591%
Tercer cuartil	4.183%	0.973%	0.829%	0.783%	0.919%	0.777%	0.785%
Máximo	128.833%	39.656%	11.557%	6.566%	46.826%	7.167%	8.509%

Tabla 27 - Datos de caja de bigotes para errores absolutos normalizados 8 a 14 neuronas.

Datos de Caja de Bigotes	8 neuronas	9 neuronas	10 neuronas	11 neuronas	12 neuronas	13 neuronas	14 neuronas
Mínimo	0.001%	0.000%	0.000%	0.000%	0.000%	0.000%	0.000%
Primer Cuartil	0.222%	0.223%	0.221%	0.228%	0.226%	0.210%	0.218%
Mediana	0.459%	0.458%	0.457%	0.456%	0.457%	0.406%	0.452%
Media	0.589%	0.581%	0.600%	0.573%	0.575%	0.549%	0.566%
Tercer cuartil	0.778%	0.776%	0.772%	0.753%	0.757%	0.683%	0.748%
Máximo	7.083%	11.625%	8.463%	7.284%	7.761%	7.803%	7.750%

Tabla 28 - Datos de caja de bigotes para errores absolutos normalizados 15 a 20 neuronas

Datos de Caja de Bigotes	15 neuronas	16 neuronas	17 neuronas	18 neuronas	19 neuronas	20 neuronas
Mínimo	0.001%	0.000%	0.001%	0.000%	0.000%	0.000%
Primer Cuartil	0.221%	0.209%	0.218%	0.207%	0.193%	0.222%
Mediana	0.457%	0.434%	0.455%	0.410%	0.408%	0.457%
Media	0.584%	0.566%	0.592%	0.528%	0.523%	0.577%
Tercer cuartil	0.784%	0.739%	0.770%	0.663%	0.674%	0.770%
Máximo	8.208%	9.236%	7.892%	10.618%	9.720%	9.450%

La red con 19 neuronas tiene un error máximo de 9.72% (Tabla 28), de forma gráfica también se muestra en la Figura 78

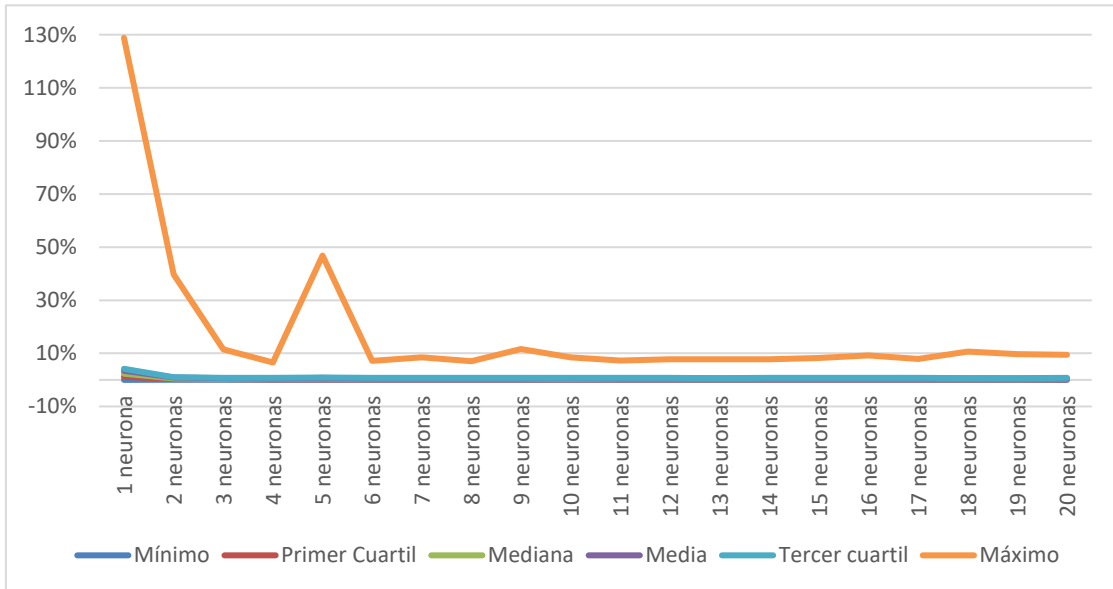


Figura 78 - Evolución de error relativo estimado de consumo de combustible para datos de test de las Tablas 23, 24 y 25

La Figura 79 muestra los valores de la Tabla 25 donde se aprecia la evolución del RMSE de entrenamiento y RMSE de test, ambos valores muestran una clara disminución a partir de la red a partir de red con dos neuronas.

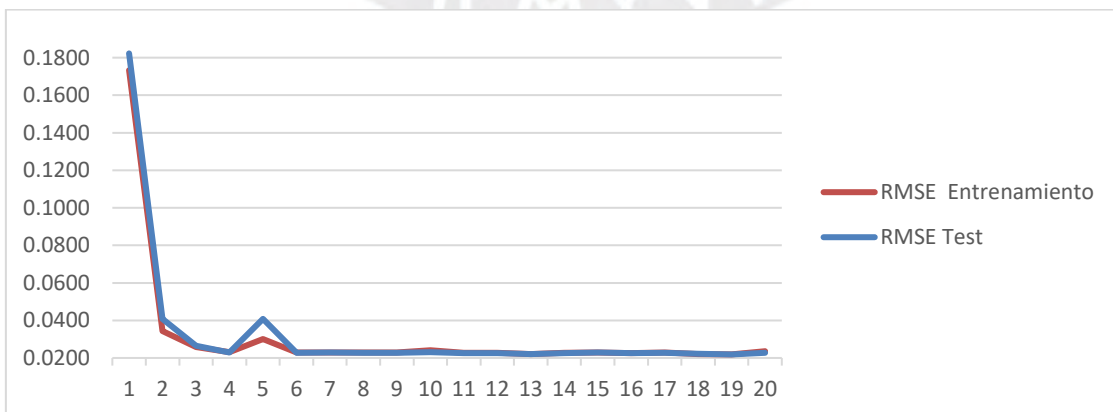


Figura 79 - RMSE para 1 neurona a 20 neuronas estimado del consumo de combustible.

EL modelo muestra la mejor validación para red con 19 neuronas, y el mejor valor se alcanza en la época 287.

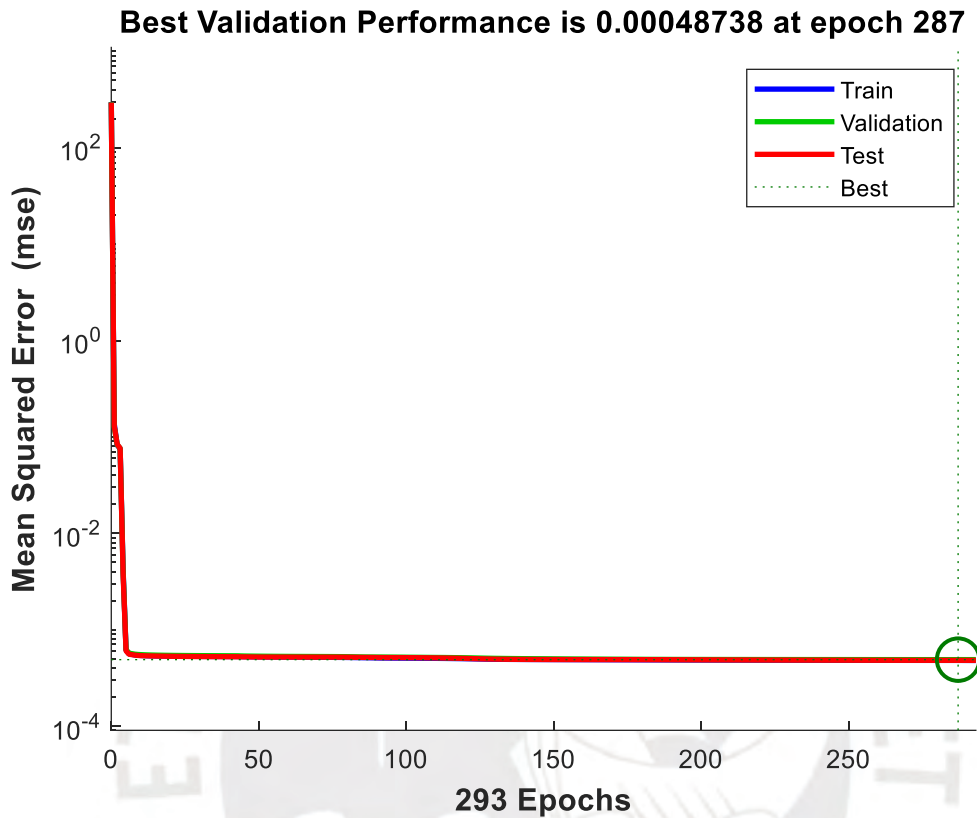


Figura 80 - MSE para entrenamiento, validación y test con 19 neuronas para estimado de consumo de gasolina

La condición de que el error no disminuye 6 veces consecutivas se cumple en la época 293, por lo que el programa realiza el paro en las iteraciones y muestra el mejor resultado seis épocas antes como se mostró (ver Figura 80 y 81).

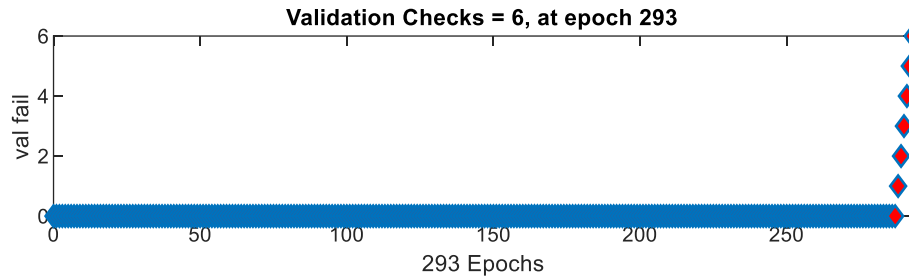


Figura 81 - Entrenamiento de la red con 17 neuronas para consumo de combustible

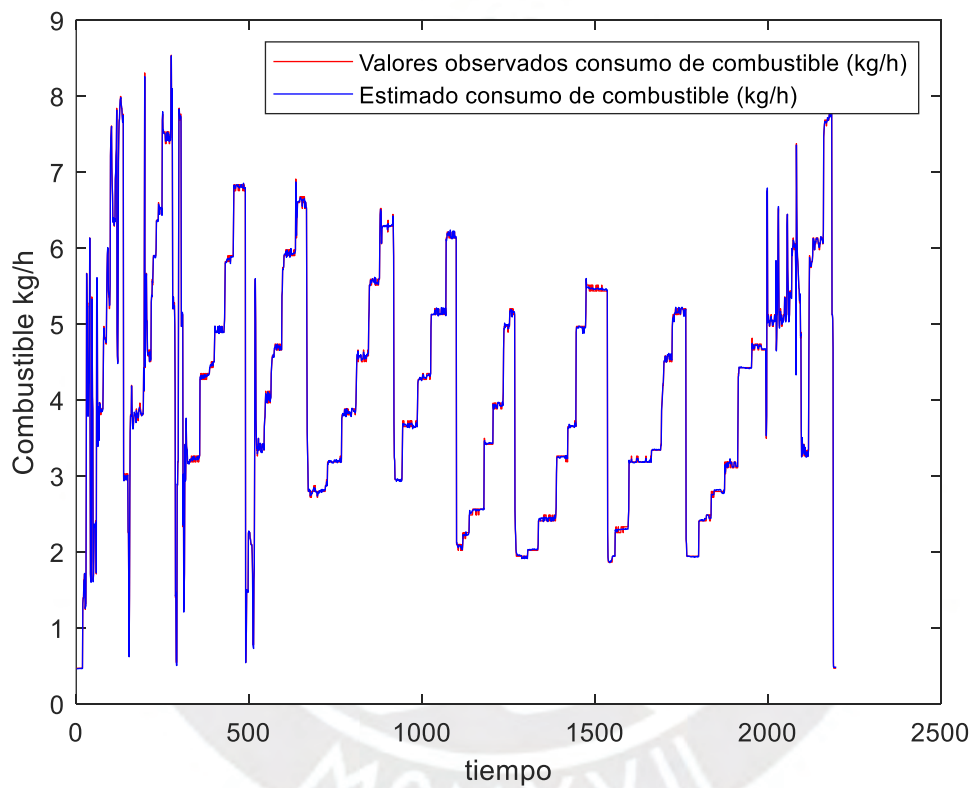


Figura 82 - Valores observados versus predicción para consumo de combustible con 19 neuronas en la capa oculta

Para consumo de combustible, los diagramas de correlación de entrenamiento y test muestran una distribución muy cercana a una recta, con pendiente 45°, lo que es indicativo de una alta correlación ver Figura 83. La Tabla 25 indica que en ambos casos el coeficiente R^2 es de 0.9998

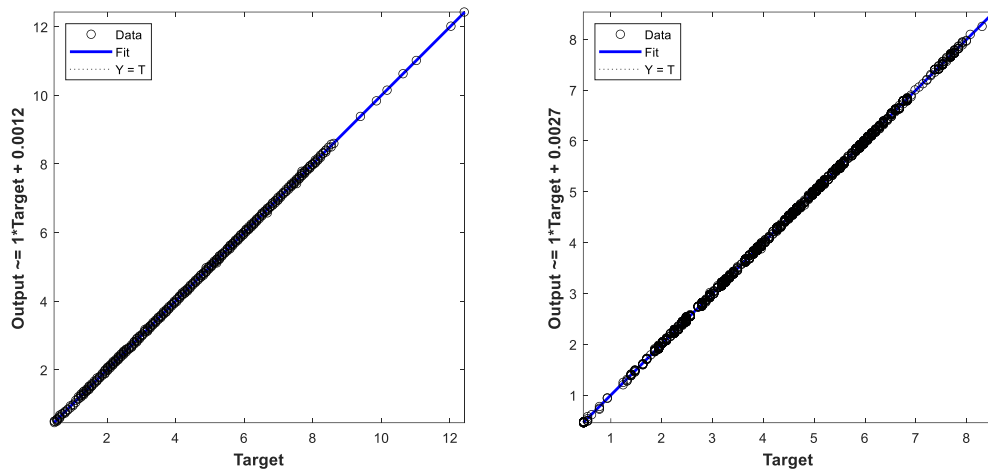


Figura 83 - Diagramas de correlación para predicción de consumo de combustible, para entrenamiento (izquierda) y test (derecha).

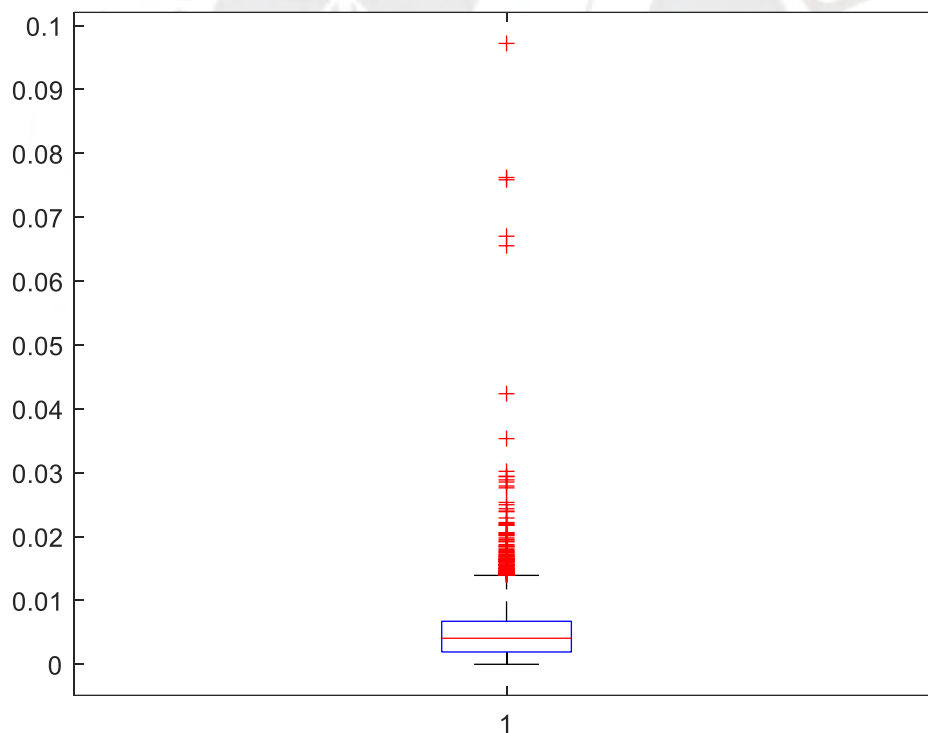


Figura 84 - Diagrama de cajas de error relativo (para test) con 113 outliers para estimado del consumo de combustible-RN con 19 neuronas en la capa oculta.

Tabla 29 - Datos de diagrama de cajas para estimado del consumo de combustible mediante Redes Neuronales de Figura 64

Mínimo	Primer Cuartil	Mediana	Media	Tercer cuartil	Máximo
0.0000	0.0019	0.0041	0.0052	0.0067	0.0972

De la Tabla 29 podemos tener que los resultados que el 75% de los datos tienen un error relativo de 0.67%, el 50% de los datos tienen un error de 0.41% y el 25% de los datos tienen un error de 0.19%.

4.3. Resumen de resultados para los modelos

A continuación, se sintetizan los resultados para estimación de CO₂ (g/s) del motor (ver Tabla 30)

Tabla 30 - Resumen de resultados para la emisión del CO₂ (g/s) estimado.

Modelo	Arreglo	Entrenamiento RMSE	Test RMSE	Tiempo de procesamiento computacional	R ² Entrenamiento	R ² Test	Error Absoluto máximo normalizado para datos de test
Regresión Lineal Múltiple	Quadratic	0.0284	0.0285	2.58	0.9996	0.9995	10.23%
Regresión Lineal Múltiple	Quadratic (Cook5)	0.0284	0.0286	6.27	0.9995	0.9995	7.35%
Máquina de Soporte Vectorial	Standardize, Kernel polinomial de orden 2.	0.0491	0.0481	8.41	0.9987	0.9987	34.69%
Redes Neuronales	13 neuronas en la capa oculta	0.0265	0.0267	6.87	0.99961	0.99960	7.85%
Redes Neuronales	13 neuronas en la capa oculta (cook5)	0.0278	0.0286	5.86	0.99961	0.99960	7.09%

De la tabla tenemos que Redes Neuronales realiza un mejor estimado para las emisiones de CO₂, con valor de RMSE (test) de 0.0267, con un error absoluto normalizado máximo para el 75% de los datos de hasta 0.99% y un

error absoluto normalizado correspondiente para él se obtiene máximo de 7.85% (logrado con un arreglo de 13 neuronas). Aun cuando los otros dos modelos tienen un coeficiente de correlación muy cercanos a 1, sus RMSE correspondiente son mayores. Para el RLM se obtiene con un RMSE de entrenamiento de 0.0284 y para RMSE de test un valor de 0.0285, con un error absoluto normalizado máximo para el 75% de hasta 1.12% y un error absoluto máximo normalizado de 10.23%, con Máquina de Soporte Vectorial se logra un RMSE de 0.0481, con un error absoluto normalizado máximo para el 75% de los datos de hasta 1.67% y un error máximo normalizado de 34.69%.

Usando la distancia de Cook, el mejor resultado se obtiene, cuando se omiten los registros que tienen una distancia de Cook mayor a 5 veces la media esta distancia. Tomando estos nuevos datos tenemos como resultado para Regresión Lineal Múltiple, (RMSE de 0.0286), con un error absoluto normalizado máximo para el 75% de hasta 1.15% y un error absoluto máximo normalizado de 7.35% menor al 10.23 % anterior, pero con mayor costo de procesamiento.

Para concluir sobre la estimación del consumo de combustible, vía los modelos, se presenta la siguiente Tabla 31:

Tabla 31 - Resumen de resultados para estimado de consumo de gasolina estimado.

Modelo	Arreglo	Entrenamiento RMSE	Test RMSE	Tiempo de procesamiento computacional	R ² Entrenamiento	R ² Test	Error Absoluto máximo normalizado para datos de test
Regresión Lineal Múltiple	Interaction	0.0230	0.0228	2.62	0.9998	0.9994	6.69%
Máquina de Soporte Vectorial	Standardize, Kernel polinomial de orden 2.	0.0611	0.0616	9.34	0.9985	0.9985	35.36%
Redes Neuronales	19 neuronas	0.0219	0.0219	14.6	0.99981	0.99981	9.72%
Redes Neuronales	4 neuronas	0.0230	0.0229	8.75	0.99979	0.99979	6.57%

Utilizando Redes Neuronales se alcanza un RMSE para test de 0.0219 frente a un RMSE de 0.0228 para el modelo RLM y un RMSE de 0.0616 para el modelo SVM (Tabla31).

De la información de la caja de bigotes concluimos que, estimando el consumo de combustible para datos de test, con Redes Neuronales se obtiene tiene 113 outliers (con un error máximo de 9.72% para 19 neuronas), con Máquina de soporte vectorial se obtiene 218 outliers (con un error máximo de 35.36%) y con Regresión Lineal Múltiple se encuentran 211 outliers (con un error máximo de 6.69%)

CONCLUSIONES

Consumo de combustible y emisiones de CO₂ para un motor de 130 HP a gasolina fueron estimados con tres modelos de regresión; Regresión Lineal Múltiple, SVM y Redes Neuronales. Para estos modelos se usó parámetros instantáneos de funcionamiento del motor (recolectados durante la fase de experimentación), esto por tratarse de modelos de escala microscópica, (modelos a escala macroscópica requieren otros datos como: masa del vehículo, coeficiente aerodinámico, entre otros) siguiendo lo empleado por Zhang y colegas [17]. En todos los modelos se obtuvo valores de RMSE, R² aceptables (muy cercanos a uno para R²), siendo las Redes Neuronales que produce estimados más cercanos a los valores observados y recolectados durante las pruebas llevadas a cabo en el laboratorio.

Un procedimiento experimental para la obtención de datos de funcionamiento de un motor a gasolina (con aplicación en vehículos livianos) instalado en un banco de pruebas fue elaborado. Para ello se aseguró el normal funcionamiento del motor, mediante un análisis de fallas por medio de la obtención de códigos DTC con un scanner conectado al puerto OBDII del módulo de control. Instalados sensores e instrumentación se efectuaron pruebas preliminares y se comprobó que registros obtenidos cumplen con los parámetros para este tipo de motor según su ficha técnica.

Datos de sensores e instrumentos entre ellos del medidor de consumo de combustible y el analizador del CO₂ en los gases de escape fueron recolectados para diferentes condiciones de torque y régimen de giro del motor. El uso de estos parámetros instantáneos de motor va de acuerdo con

el modelo de escala microscópica, como fue el caso de este trabajo. En total se recolectaron 23370 registros cada uno con cinco variables predictores (Presión y temperatura a la entrada del múltiple de admisión, régimen de giro del motor, flujo de aire al ingreso al motor y torque efectivo aplicado al motor) y dos variables objetivo (emisión de CO₂ y consumo de gasolina).

Las estimaciones a partir de los datos recolectados fueron procesadas mediante las técnicas de Regresión Múltiple, Máquina de Soporte vectorial y Redes Neuronales. Las pruebas a los modelos se realizaron con datos no utilizados en la parte de entrenamiento y validación.

Para la estimación de emisiones de CO₂, el modelo de Redes Neuronales mostró mejores resultados que Máquina de Soporte Vectorial y Regresión Lineal Múltiple. Las Redes Neuronales con trece neuronas en la capa oculta mostraron mejores resultados con un valor de RMSE de 0.0267 y un error absoluto normalizado máximo de 7.85%. El incremento de neuronas no tuvo relación directa con la disminución del RMSE, una red con 20 neuronas obtuvo un RMSE de 0.0275 (mayor al que obtuvo la Red con 13 neuronas).

Para la estimación de consumo de combustible, el modelo de Redes Neuronales y Regresión Lineal Múltiple tuvieron mejores resultados que Máquina de Soporte Vectorial: Redes Neuronales (con 19 neuronas en la capa oculta) obtuvo un RMSE de 0.0219 y Regresión Lineal Múltiple obtuvo un RMSE de 0.0228. Para el error máximo normalizado: Redes Neuronales obtuvo un valor de 9.72% y Regresión Lineal Múltiple obtuvo 6.69%. Comparando la cantidad de Outliers Redes neuronales obtuvo 113 frente a los 211 que obtuvo Regresión lineal Múltiple.

Al comparar Redes Neuronales, frente a Máquina de Soporte Vectorial y Regresión Lineal Múltiple, y se destacó diferencias en el tiempo de procesamiento, ya que con Redes Neuronales se requiere menos intervención en la preparación de datos y en el procesamiento. En el caso de la Regresión Lineal Múltiple, como por ejemplo en el caso “polyijk”, éste requiere de un análisis de correlación de las diferentes variables para asignarle exponentes, dependiendo de la importancia de la contribución en el estimado de la variable objetivo o dependiente, por parte de las variables independientes o predictoras, que en este caso son cinco.

Otra diferencia entre ellos fue el resultado del modelo, es decir lo que el modelo muestra como resultado de su entrenamiento. En el caso de la Regresión Lineal Múltiple obtuvimos hasta 25 términos o más; además se tiene el valor del coeficiente de la variable predictora, esto no sucede para el caso del modelo SVM ni para Redes Neuronales en donde el arreglo es de una cantidad de valores de pesos y sesgos tan grande que no se muestra, (y si se muestra no es posible comprender su interrelación ya que no es un arreglo polinómico) esto sumado a que el proceso de asignación inicial de pesos es aleatorio.

Una interacción entre los modelos fue realizada en el caso de estimación de emisiones de CO₂, en donde se usaron las técnicas de distancia de Cook para filtrar los registros usado para Regresión Lineal Múltiple, creando dos archivos de datos en donde se excluyó los valores en donde la distancia de Cook era 3 y 5 veces el promedio de este parámetro. Esto permitió mejorar el resultado de la técnica de Redes Neuronales, en donde se obtuvo una mejora por

disminución de outliers en de 100 a 95; y también en el error absoluto normalizado máximo de 7.85% a 7.08%.



RECOMENDACIONES

El registro de las diferentes variables debe considerar el desfase que pueda haber en el tiempo, es decir la toma de registro de consumo de combustible debe ser cercano al motor dado que de lo contrario esta podría estar desfasada en el tiempo respecto a otras variables registradas. Esta consideración debe ser la misma para otras variables.

Tener un registro doble para algunas variables como respaldo y para verificar que los datos obtenidos puedan ser contrastados antes de su análisis

Crear una base de datos de cada investigación para forme parte junto con otras investigaciones por ejemplo de motores de otras potencias o categorías y que éstas puedan ser posteriormente estudiadas.

Antes de realizar un modelo basado en datos obtenido de motor se debe asegurar que el motor se encuentre en buen estado de funcionamiento. Un motor en mal estado puede darnos un registro que al ser analizado puede que genere un modelo que estime bien, pero que al tener un mal funcionamiento del motor este modelo no puede ser utilizado.

Otra ventaja de los modelos SVM y Redes Neuronales es que pueden trabajar con variables categóricas, es decir en el caso de tener que analizar motores de diferentes años de fabricación o con otro tipo de variable categórica se puede analizar en un solo modelo de regresión incluyendo esta variable en el análisis. Esto podría ser implementado en el futuro por otro investigador.

BIBLIOGRAFÍA

- [1] M. Xionfan, H. Gang y C. Junli, «The significant roles of anthropogenic aerosols on surface temperature under carbon neutrality,» *Elsevier B.V. and Science China Press*, pp. 470-473, 2021. [DOI: 10.1016/j.scib.2021.10.022](https://doi.org/10.1016/j.scib.2021.10.022).
- [2] I. Maita, M. Morales, A. Cañari, W. Mak, A. Aquino y N. Espinoza, *CAMBIO CLIMÁTICO, CALIDAD DEL AIRE Y TRANSPORTE GUÍA PARA CUANTIFICAR LAS EMISIONES DE GEI Y CCVC EN EL SECTOR*, Lima: Ministerio de Transportes y Comunicaciones - Dirección General de Asuntos Ambientales, 2020.
- [3] «elperuano,» 2021. [En línea]. Available: <https://elperuano.pe/noticia/131529-parque-automotor-peruano-vehiculos-electricos-llegarian-al-5-en-el-2030>. [Último acceso: 14 Agosto 2023].
- [4] G. d. E. E. y Estadística, «<https://aap.org.pe>,» Asociación Automotriz del Perú, Agosto 2022. [En línea]. Available: <https://aap.org.pe/informes-estadisticos/agosto-2022/Informe-Agosto-2022.pdf>. [Último acceso: 15 Septiembre 2022].
- [5] MTC, *Anuario Estadístico 2021*, Lima: Ministerio de Transportes y Comunicaciones., 2021.
- [6] A. Velaz Acera, «Aplicación para la obtención de parámetros de conducción con el sistema OBD,» Universidad de Valladolid, Valladolid, 2019.
- [7] «<https://obd2-elm327.es/>,» [En línea]. Available: <https://obd2-elm327.es/sistema-obd2-historia-descripcion-futuro>. [Último acceso: 8 enero 2023].
- [8] MathWorks, *Statistics and Machine Learning Toolbox User's Guide*, 2021.
- [9] MathWorks, «Determine Outliers Using Cook's Distance,» de *Statistics and Machine Learning Toolbox™ User's Guide*, MathWorks, 2021, pp. 717-719.
- [10] MathWorks, «Leverage,» de *Statistics and Machine Learning Toolbox™ UserGuide*, MathWorks, 2021, p. 740.
- [11] A. KOWALCZYK, *SUPPORT VECTOR MACHINES*, Morrisville: SYNCFUSION, 2017.
- [12] Pedro Isasi Viñuela, Inés M. Galván León, «Introducción a las Redes de Neuronas Artificiales,» de *Redes de Neuronas Artificiales Un Enfoque Práctico*, Madrid, Pearson Prentice Hall, 2004, p. 227.
- [13] MathWorks,«<https://la.mathworks.com>,»MathWorks,[Enlínea].Available: https://la.mathworks.com/help/deeplearning/ref/trainlm.html?searchHighlight=levenberg&s_tid=srchtitle_levenberg_1. [Último acceso: 11 Enero 2023].
- [14] Mathworks, «<https://la.mathworks.com/>,» [En línea]. Available: https://la.mathworks.com/help/deeplearning/ref/trainbr.html?searchHighlight=trainbr&s_tid=srchtitle_trainbr_1. [Último acceso: 23 1 2023].
- [15] Martin T.Hagan;Howrad B. Demuth;Mark Hudson Beale; Orlando De Jesús, «NEURAL NETWORK DESING (2nd Edición),» 2014. [En línea]. Available: <https://www.onlineprogrammingbooks.com/neural-network-design-2nd-edition/>.

[Último acceso: 26 12 2022].

- [16] MathWorks, «Coefficient of Determination (R-Squared),» de *Statistics and Machine Learning Toolbox™ User's Guide*, MathWorks, 2021, pp. 723-724.
- [17] R. Zhang, Y. Wang, Y. Pang, B. Zhang, Y. Wei, M. Wang y R. Zhu, «A Deep Learning Micro-Scale Model to Estimate the CO₂ Emissions from Light-Duty Diesel Trucks Based on Real-World Driving,» *atmosphere*, vol. 13, nº 1466, 2022. [DOI: 10.3390/atmos13091466](https://doi.org/10.3390/atmos13091466).
- [18] M. Madziel, A. Jaworski, H. Kuszewski, P. Woś, T. Campisi y K. Lew, «The Development of CO₂ Instantaneous Emission Model of Full Hybrid Vehicle with the Use of Machine Learning Techniques.,» *energies*, vol. 15, nº 142, 2022. [DOI: 10.3390/en15010142](https://doi.org/10.3390/en15010142).
- [19] M.-C. Raquel Modet, «MODELOS DE “MACHINE LEARNING” PARA LA ESTIMACIÓN DEL CONSUMO DE COMBUSTIBLE EN VEHÍCULOS,» Escuela Técnica Superior de Madrid - Universidad Politécnica de Madrid, Madrid, 2021.
- [20] E. Martín Gutiérrez, «APLICACIÓN DE MODELOS DE REDES NEURONALES RECURRENTE A LA PREDICCIÓN DE EMISIONES CONTAMINANTES DE AUTOBUSES URBANOS,» UNIVERSIDAD POLITÉCNICA DE MADRID. ESCUELA TÉCNICA SUPERIOR DE INGENIEROS INDUSTRIALES, Madrid, 2019.
- [21] OMS, «<https://www.who.int/>,» Organización Mundial de la Salud, [En línea]. Available: <https://www.who.int/teams/environment-climate-change-and-health/air-quality-and-health/health-impacts/types-of-pollutants>. [Último acceso: 1 Diciembre 2022].
- [22] A. Nova Garcia, «APLICACIÓN DE MODELOS DE REDES NEURONALES AL MODELADO Y PREDICCIÓN DE EMISIONES DE AUTOBUSES URBANOS. Escuela Técnica Superior de Ingenieros Industriales,» UNIVERSIDAD POLITÉCNICA DE MADRID, Madrid, 2018.
- [23] L. Qing, Q. Fengxiang y Y. Lei, «A Machine Learning Approach for Light-Duty Vehicle Idling Emission Estimation Based on Real Driving and Environmental Information,» *Environ Pollut Climate Change*, vol. 1, nº 106, p. 7, 2016. [DOI:10.4172/2573-458X.1000106](https://doi.org/10.4172/2573-458X.1000106)
- [24] J. Y. Rosales Malpartida, «Predicción de la contaminación atmosférica generada por las emisiones del CO₂ en el Perú utilizando los métodos ARIMA y Redes Neuronales,» *Tecno Humanismo*, vol. 2, nº 1, pp. 114-125, 2022.
- [25] E. F. Méndez Garcés y J. A. Herrera Burgos, «Sistema basado en redes neuronales artificiales orientado a la predicción del dióxido de carbono (CO₂) como índice de contaminación, en la zona céntrica de la ciudad de Santo Domingo.,» UNIVERSIDAD REGIONAL AUTÓNOMA DE LOS ANDES, SANTO DOMINGO -ECUADOR, 2019.
- [26] R. A. Ortiz Garcia y F. Ceballos, «Análisis de datos para la optimización de la gestión de flotas vehiculares: Impacto en los costos operativos y rendimiento empresarial,» Universidad de Antioquia, Medellín, Antioquia, Colombia, 2023.

- [27] «OEM-Specific Enhanced Diagnostics Support Coverage,» 2019. [En línea]. Available: https://www.obdlink.com/wp-content/uploads/2019/12/oem-specific_coverage.pdf. [Último acceso: 27 10 2023].
- [28] D. Ajtay, M. Weilenmann y P. Soltic, «Towards accurate instantaneous emission models,» *Atmospheric Environment*, vol. 39, nº 13, pp. 2443-2449, 2005. DOI: [10.1016/j.atmosenv.2004.03.080](https://doi.org/10.1016/j.atmosenv.2004.03.080)
- [29] Willian W. Hines, Douglas C Montgomery, «Diagrama de Caja,» de *Probabilidad y estadística para ingeniería y administración*, México, Compañía Editorial Continental, S.A. de C.V. México, 1996, p. 23.
- [30] MathWorks, «Choose a Model or Range of Models,» de *Statistics and Machine Learning Toolbox™ User's Guide*, MathWorks, 2021, p. 673.
- [31] G. Wilkinson y C. Rogers, «Symbolic Description of Factorial Models for Analysis of Variance,» *Journal of the Royal Statistical Society. Series C (Applied Statistics)*, vol. 22, nº 3, pp. 392-399, 1973.
- [32] H. Jae Hyuk, «Comparing Models for Time Series Analysis.,» University of Pennsylvania. Wharton Research Scholars, 2018.
- [33] P. Kim, *MATLAB Deep Learning With Machine Learning, Neural Networks and Artificial Intelligence*, Soeul: Apress, 2017.

Anexo A

Comandos de propiedades de Red Neuronal tomado de [8]. Los comandos se escriben seguidos del nombre de la red y unidos por un punto.

Tabla A.1 - Comandos de Red Neuronal. Matlab [8].

numInputs	Número de entradas
numLayers	Número de capas
biasConnect	Número de capas por un vector
inputConnect	Número de capas por número de entradas
layerConnect	Numero de capas por número de capas
outputConnect	Uno por número de capas



Anexo B

Tabla B.1- Características de motor Hyundai (proporcionado por el Laboratorio de Energía de la PUCP).



Screenshot of a vehicle information system showing details for a red Hyundai car. The VIN is KMHDG41CAEU295628 and the manufacturing date is 24/07/2014. The main attributes section is active, showing dropdown menus for body type, engine capacity, engine, fuel, transmission, drive type, and climate type.

Atributos principales	General	Códigos de opción
CARROCERIA	(S4) SEDAN - 4DR 5P	
CAPACIDAD DEL MOTOR	(D2) 1600 CC - GAMMA	
MOTOR	(6) DOHC - MPI	
COMBUSTIBLE	(1) GASOLINE - UNLEADED	
CAJA DE CAMBIOS	(7) MANUAL T/M - 6 SPEED 2WD	
MANIVELA	(L) DRIVE TYPE - LHD	
TIPO DE CLIMA	(1) WEATHER - THE TROPICAL ZONE	

Anexo C

Tabla C.1 - Ficha técnica de combustible utilizado. Fuente: Petroperú.

CLASE DE PRODUCTO		COMBUSTIBLE		Fecha efectiva: 01.01.2023	
TIPO DE PRODUCTO		COMBUSTIBLE USO MOTOR		Reemplaza edición de: -	
NOMBRE DE PRODUCTO					
GASOLINA PREMIUM					
ENSAYOS	ESPECIFICACIONES (a)		MÉTODO		
	MÍN.	MÁX.	ASTM	OTROS	
APARIENCIA	Claro/Brillante, libre de agua y partículas				
Color comercial	Amarillo (b)				
VOLATILIDAD					
Gravedad API a 60°F	Reportar		D-1298, D-4052		
Densidad a 60°F, g/cc	Reportar		D-1298, D-4052		
Destilación, a 760 mm Hg, °C			D-86, D-7096		ISO-3405
Punto inicial de ebullición	Reportar				
5 %V recuperado	Reportar				
10 %V recuperado	65				
20 %V recuperado	Reportar				
50 %V recuperado	77	118			
90 %V recuperado	190				
95 %V recuperado	Reportar				
Punto final de ebullición	225				
Recuperado, %V	Reportar				
Residuo, %V	2.0				
Pérdida, %V	Reportar				
Relación vapor/líquido = 20, °C	47		D-5188, D-4814		
Presión de vapor, psi (KPa)	10 (69)		D-323, D-4953, D-5191, D-6378		ISO 3007, UNE EN 13016-1
Índice de manejabilidad	630		D-4814		
CORROSIVIDAD					
Corrosión lámina de cobre, 3h, 50°C, N°	1		D-130		ISO 2160
Azufre total, mg/Kg	50		D-2622, D-3120, D-5453, D-6920, D-7039, D-7220		ISO-13032, ISO 20846, ISO 20884, ISO 8754
ANTIDETONANCIA					
Número de octano Research	95.0		D-2699		ISO 5164
COMPOSICIÓN					
Aromáticos, % Vol.	45		D-1319, D-5580, D-6839, D-6730		ISO 22854, UNE EN 15553
Olefinas, % Vol.	25		D-1319, D-6839, D6730		ISO 22854, UNE EN 15553
Benceno, % Vol.	2		D-3606, D-5580, D-6839, D6730		ISO 22854, UNE EN 238, UNE EN 12177
Oxígeno, % masa	Reportar		D-4814, D-4815, D5599, D-5845, D-6730		ISO 22854, UNE EN 13132
ESTABILIDAD A LA OXIDACIÓN					
Período de inducción, minutos	240		D-525		ISO 7536, UNE 51203
CONTAMINANTES					
Gomas lavadas, mg/100mL	5.0		D-381		ISO 6246
Plomo, g/L	0.013		D-3237, D-5059		UNE EN 237, UNE EN 13723
Manganeso, mg/L	0.25		D-3831		UNE EN 16136
OBSERVACIONES:					
(a) En concordancia con la Resolución Ministerial N° 469-2021-MINEM/DM.					
(b) No contiene colorante, corresponde al color natural de la gasolina.					

Anexo D

Tabla D.1 - Certificado de calibración del analizador de gases de escape vehicular - CC-0443-

22. Página 1 de 3.



**LABORATORIO DE CALIBRACIÓN ACREDITADO POR
EL ORGANISMO PERUANO DE ACREDITACIÓN
INACAL-DA CON REGISTRO N°LC-049**



Formulario CALIBRA R.C-017 N08
Página 1 de 3

CERTIFICADO DE CALIBRACIÓN CC-0443-22

INFORMACION DEL SOLICITANTE

Solicitud de Servicio	SS-1469-22
Nombre	PONTIFICIA UNIVERSIDAD CATOLICA DEL PERU
RUC	20155945860
Dirección	AV. UNIVERSITARIA N. 1801 URB. PANDO SAN MIGUEL - SAN MIGUEL
Contacto	JULIO CESAR CUISANO IGUSQUIZA
Correo	jcuisano@pucp.edu.pe
Teléfono	985000079

INFORMACION DEL EQUIPO/INSTRUMENTO CALIBRADO

Equipo/Instrumento	Analizador de Gases Vehicular
Marca	Infrared Industries
Modelo	FGA4500
N° Serie / Identificación	6217
Clase de exactitud	OIML Clase 1
Resolución	CO: 0,01% vol; CO ₂ : 0,01% vol; HC: 1 ppm; O ₂ : 0,01 %vol
Intervalo de indicación	CO: (0 a 10,00) %vol; CO ₂ : (0 a 20) %vol; HC: (0 a 10000) ppm; O ₂ : (0 a 25) %vol

LUGAR Y FECHA DE CALIBRACIÓN

Lugar de calibración	Laboratorio de Calibración – Calibra S.A.C., Jr. San Lorenzo N°363 Surquillo, Lima, Lima
Fecha de calibración	2022-10-03

Este certificado de calibración documenta la trazabilidad a los patrones nacionales e internacionales, que realizan las unidades de medida de acuerdo con el Sistema Internacional de Unidades (SI).

Los resultados consignados en el presente Certificado de Calibración se refieren únicamente al objeto sometido a calibración; al momento y condiciones en las que se realizaron las mediciones, CALIBRA S.A.C. no se responsabiliza por los prejuicios que puedan derivarse del uso inadecuado de los objetos calibrados o del presente Certificado de Calibración.

Este certificado de calibración, al estar en el marco de la acreditación del INACAL – DA, cumple con los requisitos de la Norma Técnica Peruana NTP- ISO/IEC 17025 "Requisitos Generales para la Competencia de los Laboratorios de Ensayo y Calibración" y se encuentra dentro del ámbito de reconocimiento multilateral/multinacional de los miembros firmantes de Cooperación Interamericana de Acreditación IAAC y de la Cooperación Internacional de Acreditación de Laboratorios (ILAC).

Este certificado de calibración sólo puede ser otorgado completamente y sin modificaciones. Las modificaciones requieren la autorización de CALIBRA S.A.C. Certificadas sin firma y sello carecen de validez.

Las enmiendas al presente documento no efectuadas por CALIBRA S.A.C., constituyen un delito contra la fe pública y el infractor es sujeto de sanciones civiles y penales reguladas por dispositivos legales vigentes.

<p>Fecha de emisión</p>  <p>2022-10-04</p>	<p>Jefe de Laboratorio de Calibración</p>  <p>NELSON NARIÑEZ</p>	<p>Técnico de Calibración</p>  <p>BERNIE CONTRERAS</p>
---	---	---

CALIBRA S.A.C. – Laboratorio de Calibración - Jr. San Lorenzo N°363, Surquillo, Lima, Lima
 Teléfono: 2424636 - Celular: 994 2360 07 - Correo Electrónico: atencioncliente1@calibra.pe - Página Web: www.calibra.pe

Tabla D.2 - Certificado de calibración del analizador de gases de escape vehicular - CC-0443-

22. Página 2 de 3



Número de Certificado CC-0443-22

Página 2 de 3

METODO DE CALIBRACIÓN

La calibración se ha realizado por comparación directa con patrones, determinándose el error de indicación del analizador de gases, aplicando el procedimiento PLC-003 - "Procedimiento de calibración de instrumentos de medición de gases de escape de vehículos" Versión 03, tomando como referencia la Norma Metroológica Peruana NMP-024:2017 - "Instrumentos de Medición de Gases de Escape de Vehículos. Parte 2 Controles Metroológicos y Ensayos de funcionamiento" Sección 8.3.

CONDICIONES DE CALIBRACIÓN

MAGNITUD	MÁXIMO	MÍNIMO
Temperatura [°C]	19,5	19,2
Humedad Relativa [%]	66,7	63,9

MAGNITUD	INICIAL	FINAL
Presión [hPa]	1004	1004
Tensión [V]	221,1	220,8
Frecuencia [Hz]	59,9	59,9

PATRONES UTILIZADOS

PATRONES	CERTIFICADO DE ANÁLISIS/CALIBRACIÓN	TRAZABILIDAD METROLÓGICA
Gases Patrón	CC727076, CC727060, CC737563, EB0034254	Patrón de medición internacional del NIST
Termohigrómetro	T-3154-2020	Patrón de medición nacional del INACAL-DM
Barómetro/Medidor de Presión	LFP-269-2020	Patrón de medición internacional del CEI/MAM
Multímetro	E-0266-2020	Patrón de medición internacional del NIST

INDICACIÓN DEL EQUIPO ANTES DEL AJUSTE

COMPONENTE	VALOR PATRÓN	LECTURA	ERROR ABSOLUTO	ERROR RELATIVO [%]	AJUSTE
CO [%Vol]	3,49	3,61	0,12	3,44	SI
CO ₂ [%Vol]	14	13,29	-0,71	-5,07	SI
Hexano [ppm]	1023	1020	-3	-0,34	SI
Óxigeno [%Vol]	20,9	20,77	-0,13	-0,62	SI



LABORATORIO DE CALIBRACIÓN ACREDITADO POR
EL ORGANISMO PERUANO DE ACREDITACIÓN
INACAL-DA CON REGISTRO N°LC-049



Número de Certificado CC-0443-22

Página 3 de 3

RESULTADOS DE LA CALIBRACIÓN

MEZCLA DE GAS E	VALOR PATRÓN	LECTURA PROMEDIO	ERROR ABSOLUTO	ERROR RELATIVO [%]	INCERTIDUMBRE
CO [%Vol]	0,4988	0,49	-0,0088	-1,8	0,0076
CO ₂ [%Vol]	5,975	5,857	-0,118	-2,0	0,062
Hexano [ppm]	102	98	-4	-3,9	2,1
Oxígeno [%Vol]	0,498	0,837	0,34	68,2	0,025

PEF (200 ppm) = 0.512

MEZCLA DE GAS F	VALOR PATRÓN	LECTURA PROMEDIO	ERROR ABSOLUTO	ERROR RELATIVO [%]	INCERTIDUMBRE
CO [%Vol]	0,999	0,973	-0,026	-2,6	0,013
CO ₂ [%Vol]	10,02	9,85	-0,17	-1,7	0,1
Hexano [ppm]	306,5	296	-10,5	-3,4	6
Oxígeno [%Vol]	9,96	10,36	0,4	4,0	0,1

PEF (600 ppm) = 0.512

MEZCLA DE GAS G	VALOR PATRÓN	LECTURA PROMEDIO	ERROR ABSOLUTO	ERROR RELATIVO [%]	INCERTIDUMBRE
CO [%Vol]	3,49	3,403	-0,087	-2,5	0,036
CO ₂ [%Vol]	14	13,89	-0,11	-0,8	0,14
Hexano [ppm]	1023	989	-34	-3,4	20
Oxígeno [%Vol]	20,9	21,36	0,46	2,2	0,21

PEF (2000 ppm) = 0.512

NOTAS

- Los gases de referencia fueron suministrados a la sonda a presión ambiente (con una desviación máxima de 8 hPa).
- La incertidumbre de calibración corresponde a un factor de cobertura $k=2$ y un nivel de confianza de aproximadamente 95%.
- Con fines de identificación se ha colocado en el equipo una etiqueta autoadhesiva de color gris plata.
- El usuario es responsable de la calibración de sus instrumentos a intervalos apropiados de acuerdo a su uso, conservación, mantenimiento y las disposiciones legales vigentes.
- Las lecturas de HC son reportadas como hexano.
- El valor del oxígeno no desciende a menos de 0,1 % Vol en un periodo de 60 segundos.