


# Modelo de Temperatura de escape en el catalizador de un vehículo ICE

En esta práctica veremos cómo predecir la temperatura de escape en función de otras variables. Para ello, necesitaremos cargar un dataset con la siguiente información:


	Variable	Unidades	Descripción
0	nmot_w	rpm	Velocidad angular del Motor
1	mshfm_w	g/s	Flujo de Aire
2	rl_w	%	Carga del Motor Actual
3	rlsol_w	%	Carga del Motor Especificada
4	pvdkds_w	mbar Absolutos	Presión de Soplado del Turbo Actual
5	plsol_w	mbar Absolutos	Presión de Soplado del Turbo Especificada
6	lamsoni_w	Adimensional	Lambda Actual
7	lamsbg_w	Adimensional	Lambda Especificada
8	tikatm_w	Grados C	Temp Gas Escape: Dentro del Catalizador

Arrastra aquí tus logs en formato .CSV



Drag and drop file here  
Limit 200MB per file • CSV

Browse files

 logs\_me7logger.csv 135.5KB ×

Archivo subido correctamente.

Vista previa del archivo (tikatm\_w se ha pasado a Kelvin para ser más riguroso con los cálculos):

	nmot_w	mshfm_w	rl_w	rlsol_w	pvdkds_w	plsol_w	lamsoni_w	lamsbg_w	tikatm_w
0	2,909.25	20.5	36.6797	37.3594	1,073.59	1,010.82	0.9937	1	908.207
1	2,905.75	20.8889	37.0313	38.3438	1,073.59	1,010.82	0.9885	1	908.207
2	2,915.75	20.9445	37.4063	39.3281	1,076.64	1,010.82	0.9949	1	908.207
3	2,913	21.5278	37.875	40.4766	1,079.65	1,010.82	0.9973	1	908.012
4	2,910	21.8889	38.4141	43.3125	1,079.65	1,010.78	0.9985	1	908.012

# EDA inicial:

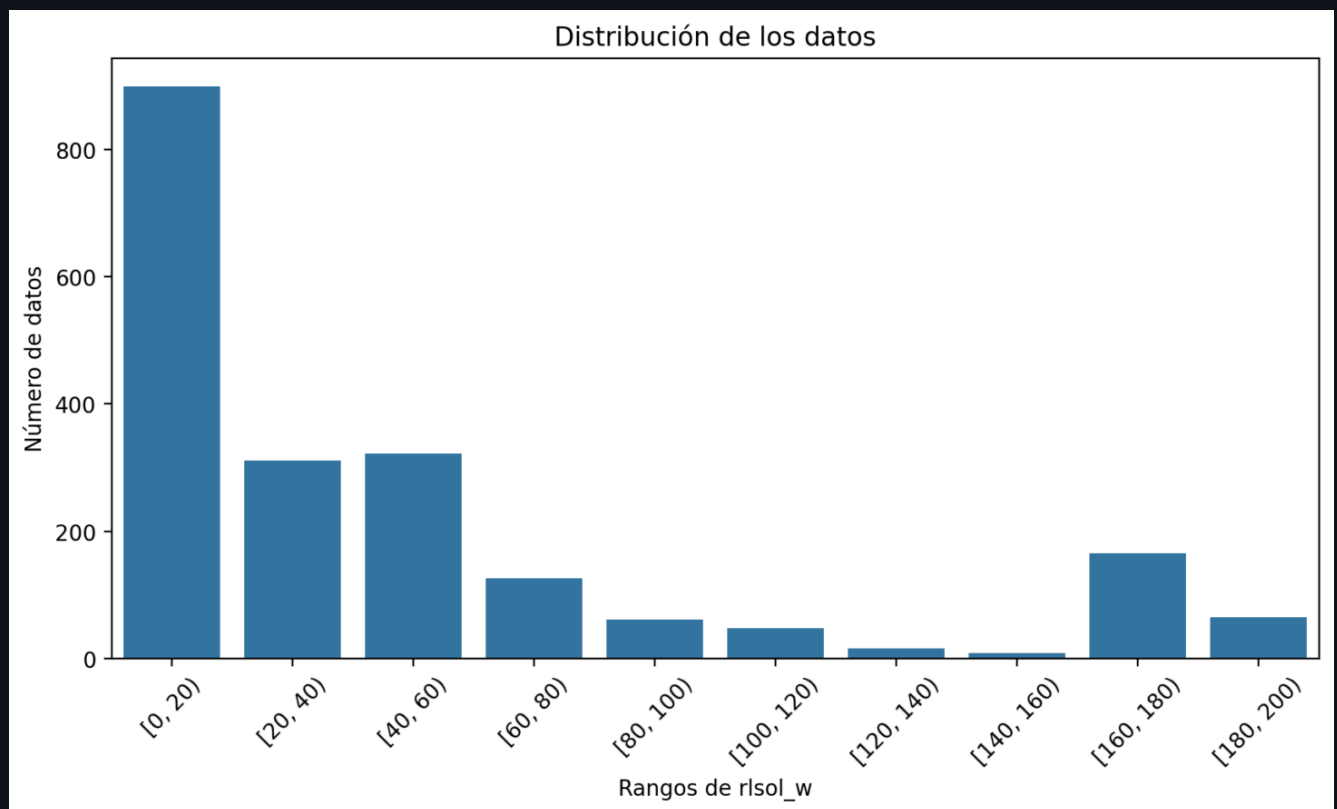
Hacemos un pequeño estudio de los datos aportados

	nmot_w	mshfm_w	rl_w	rlsol_w	pvdkds_w	plsol_w	lamsoni_w	lamsbg_w	tikatm_w
count	2,020	2,020	2,020	2,020	2,020	2,020	2,020	2,020	2,020
mean	2,246.0903	26.5052	45.0806	49.1922	1,199.0104	1,182.3358	1.522	0.9797	989.0107
std	1,292.7671	43.2384	44.5148	52.0162	326.4367	404.542	1.1736	0.0747	57.5989
min	686.75	0	10.9453	12.75	1,021.99	1,005.23	0.75	0.7969	906.273
25%	1,500.1875	4.3611	13.2891	14.25	1,036.3675	1,007.46	0.9512	1	951.722
50%	2,020.25	6.8472	23.9883	23.332	1,049.34	1,009.38	0.9973	1	969.281
75%	2,323.1875	22.6389	57.2754	54.6856	1,149.53	1,009.96	1.044	1	994.769
max	6,478.5	159.695	176.297	185.414	2,358.75	2,349.53	4	1.0469	1,119.066

Correlación entre variables

nmot_w	1.00	0.82	0.66	0.68	0.75	0.74	0.09	-0.52	0.52
mshfm_w	0.82	1.00	0.93	0.94	0.95	0.96	-0.26	-0.71	0.19
rl_w	0.66	0.93	1.00	0.98	0.95	0.92	-0.37	-0.60	0.02
rlsol_w	0.68	0.94	0.98	1.00	0.93	0.95	-0.35	-0.63	0.05
pvdks_w	0.75	0.95	0.95	0.93	1.00	0.95	-0.24	-0.64	0.09
plsol_w	0.74	0.96	0.92	0.95	0.95	1.00	-0.24	-0.67	0.11
lamsoni_w	0.09	-0.26	-0.37	-0.35	-0.24	-0.24	1.00	0.37	0.41
lamsbg_w	-0.52	-0.71	-0.60	-0.63	-0.64	-0.67	0.37	1.00	-0.11
tikatm_w	0.52	0.19	0.02	0.05	0.09	0.11	0.41	-0.11	1.00
	nmot_w	mshfm_w	rl_w	rlsol_w	pvdks_w	plsol_w	lamsoni_w	lamsbg_w	tikatm_w

Puede que los datos estén desbalanceados. Aquí lo veremos en función de la carga (Variable primaria de la bosch me7.5):



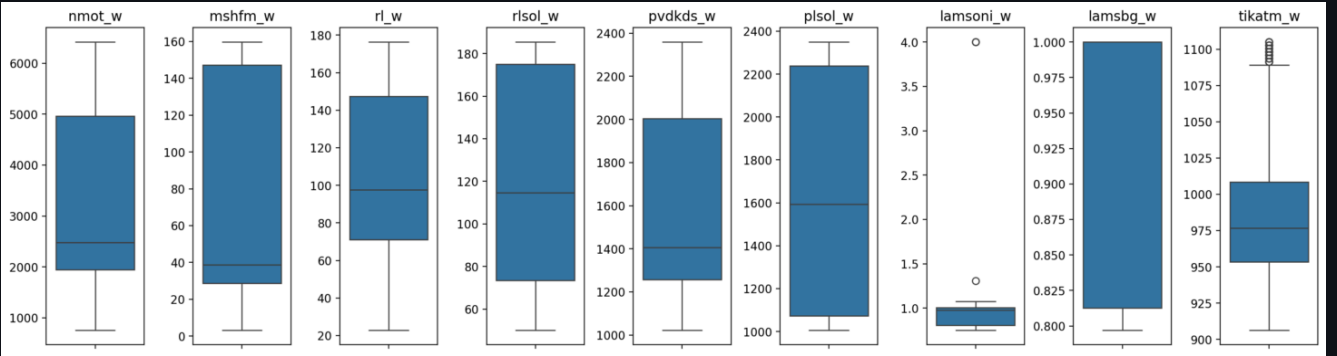
## EDA Dinámico:

Selecciona el rango de valores a estudiar



	nmot_w	mshfm_w	rl_w	rlsol_w	pvdks_w	plsol_w	lamsoni_w	lamsbg_w	tikatm_w
count	522	522	522	522	522	522	522	522	522
mean	3,334.1221	79.5829	108.82	124.4263	1,606.3755	1,682.2945	0.9309	0.9241	986.6284
std	1,655.3288	57.2562	41.0904	49.2885	425.7939	544.5034	0.2145	0.0929	47.8663
min	755	3.1389	22.8047	50.0156	1,021.99	1,005.55	0.75	0.7969	906.273
25%	1,946	28.7778	71.045	73.4473	1,256.6525	1,072.35	0.8084	0.8125	953.5143
50%	2,480.875	38.8611	97.6875	114.715	1,406.27	1,593.495	0.9767	1	976.732
75%	4,958.125	147.25	147.51	175.0785	2,003.31	2,239.47	1.0047	1	1,008.431
max	6,414.75	159.695	176.297	185.414	2,358.75	2,349.53	4	1	1,104.906

Boxplots de todas las variables con los datos acotados:



# Modelos

A continuación se entrenarán 3 modelos con los datos acotados: Regresión lineal, Puntos interpolados linealmente y XGBoost.

Es esperable que obtengamos los mejores resultados con XGBoost, pero intentaremos alcanzar buenos resultados con los 2 primeros para que su implementación en microcontroladores y ejecución en tiempo real sea viable.

Su entrenamiento tendrá lugar en combinación con validaciones cruzadas y búsqueda de los mejores hiperparámetros para adecuarse lo mejor posible al rango seleccionado arriba.

El entrenamiento comenzará cuando pulses el botón. Por favor, ten paciencia, puede tardar entre 2 y 5 minutos dependiendo de tu HW.

Entrenar Modelos

Con la Regresión lineal hemos obtenido las siguientes métricas de error:

$R^2$ : 0.8677

MSE: 302.49°C<sup>2</sup>

Error medio: 17.39°C

Con la Multipunto + interpolación lineal estos son los mejores resultados:

Mejores hiperparámetros: {'spline\_\_degree': 1, 'spline\_\_n\_knots': 9}

$R^2$ : 0.9754

MSE: 53.60°C<sup>2</sup>

Error medio: 7.32°C

Con el XGBoost esto son los mejores resultados:

Mejores hiperparámetros: {'learning\_rate': 0.25, 'max\_depth': 6, 'n\_estimators': 150}

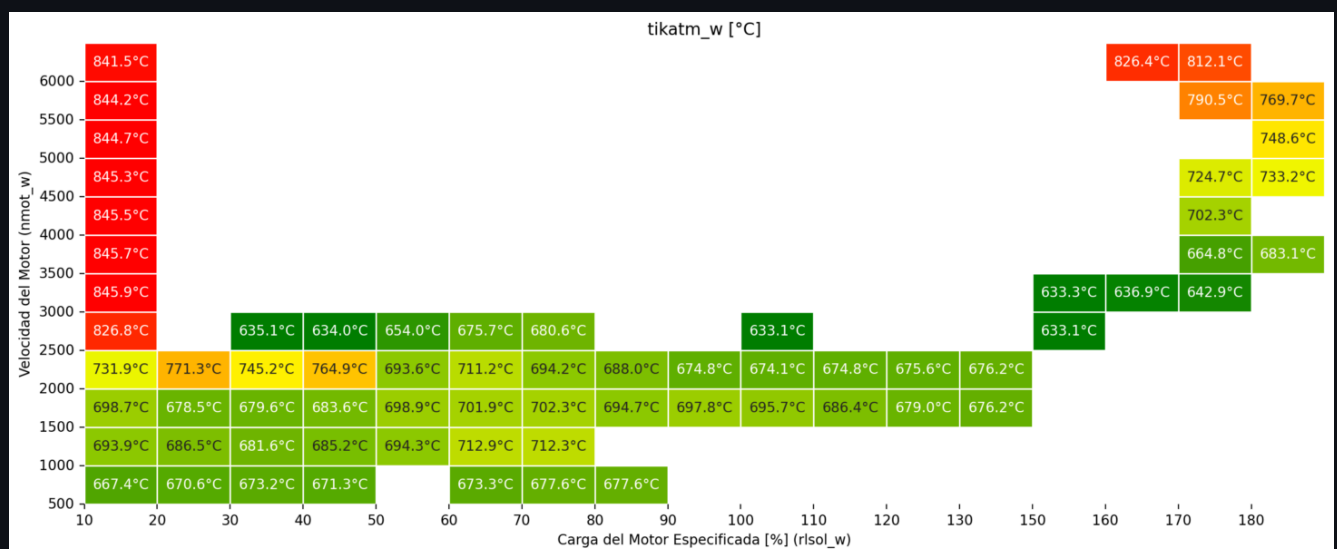
$R^2$ : 0.9904

MSE: 21.93°C<sup>2</sup>

Error medio: 4.68°C

## Puesta en contexto

tikatk\_w en f(nmot y rlsol\_w)



Las celdas rojas en cargas bajas aparentemente no tienen sentido.

Por qué ese cambio tan abrupto en 2500rpm? Y por qué no hay muestras para esas rpm en la segunda columna de carga?

Los ensayos que típicamente se hacen cuando se modifica el SW de control de motor de un coche, consisten en lanzadas en 3ª marcha con el acelerador a fondo desde unas 2500rpm hasta la zona más alta del tacómetro.

Cuando se llega, se suelta el acelerador y se deja el coche reducir de velocidad en punto muerto para medir las pérdidas mecánicas de potencia.

Con lo cual, la carga solicitada va a bajar repentinamente y las rpm van a disminuir rápidamente a ralentí, pero viniendo de una situación de estrés máximo y las temperaturas del gas y elementos de escape seguirán siendo muy altas.

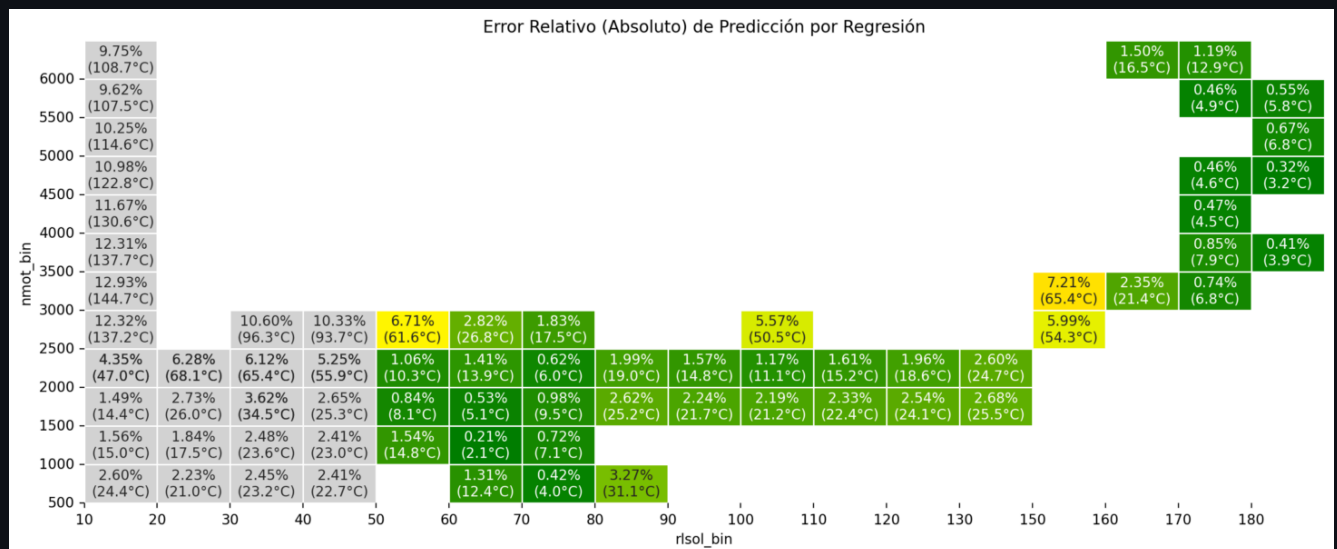
Entre 500 y 2500rpm ya apreciamos temperaturas más moderadas debido a que hay mucho dato de crucero.

Para hacer un modelo de temperatura de escape, deberían haberse contemplado más variables y hacer ensayos manteniendo cada variable fija mientras varía el resto, para tener toda la zona del gráfico completa y con una densidad homogénea de datos

Pero para este proyecto, tiraremos con el dataset que tenemos.

## Errores de Regresión

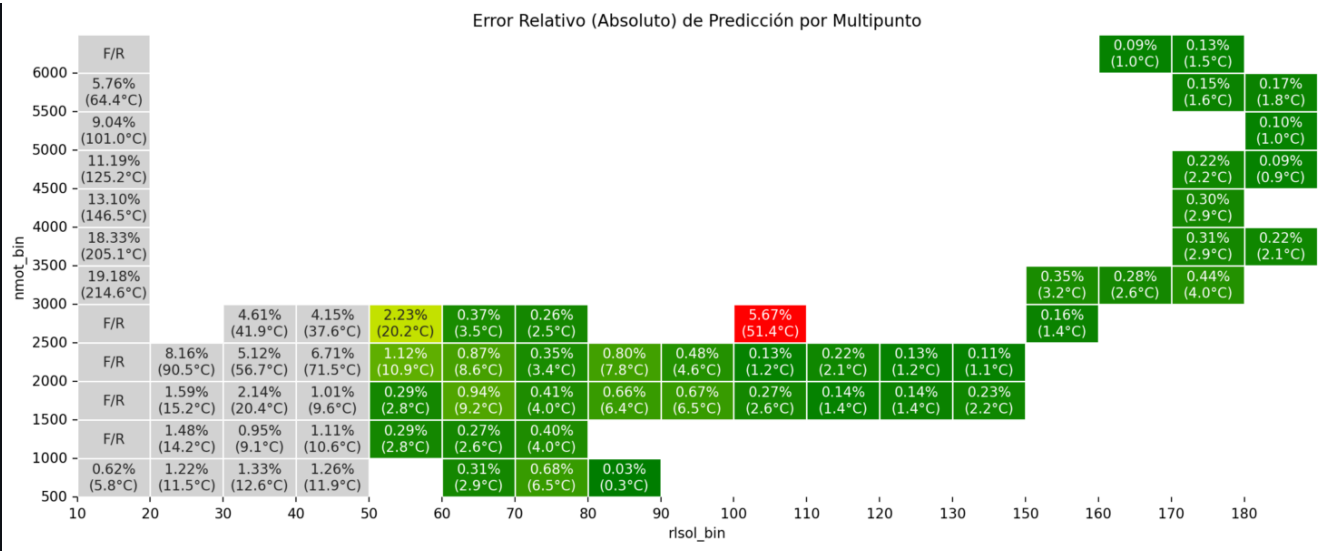
En gris: Predicciones fuera del área de entrenamiento



## Errores de Multipunto

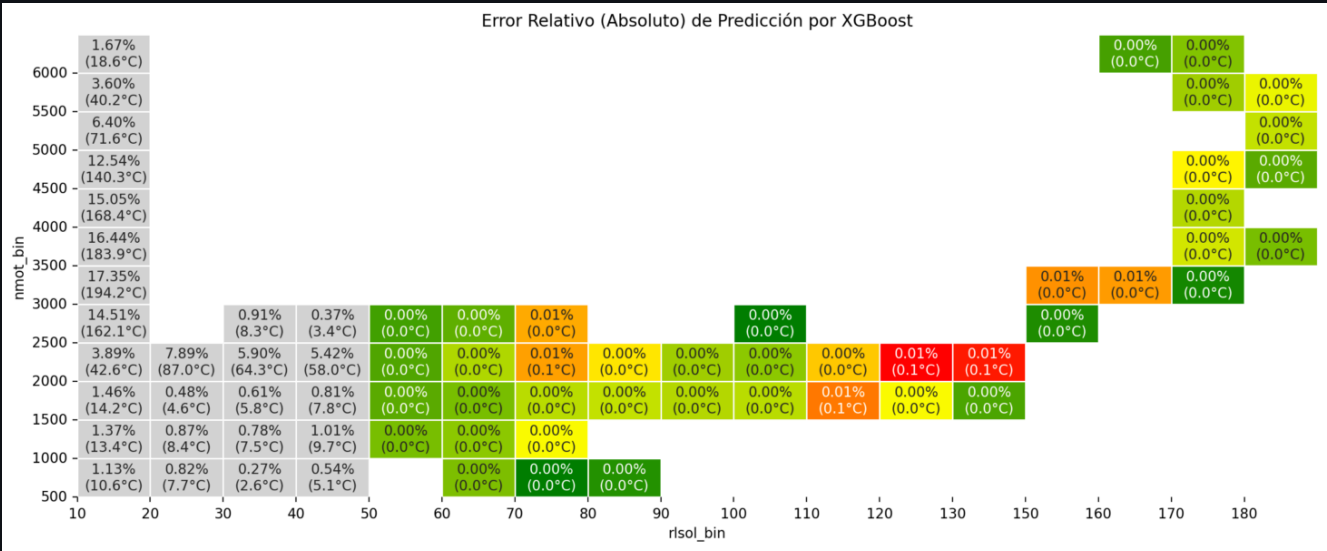
En gris: Predicciones fuera del área de entrenamiento

F/R = Fuera de rango



# Errores de XGBoost

En gris: Predicciones fuera del área de entrenamiento



Entrenamiento completado