

Detección de Fallos Utilizando Deep Learning

Análisis de Datos de Vibraciones para un Diagnóstico Preciso de Fallos

Ávalos, Silvio
Orzusa, Emanuel
Parodi, Giulio

22 de junio de 2024

Introducción

- ▶ Importancia de la detección de fallos en entornos industriales.
- ▶ Breve introducción al conjunto de datos: 39 escenarios con 25 pruebas cada uno, total 975 archivos CSV por cada RPM.
- ▶ Conjunto de datos de escenarios de fallas simples y dobles utilizando señales de vibración de una máquina rotativa

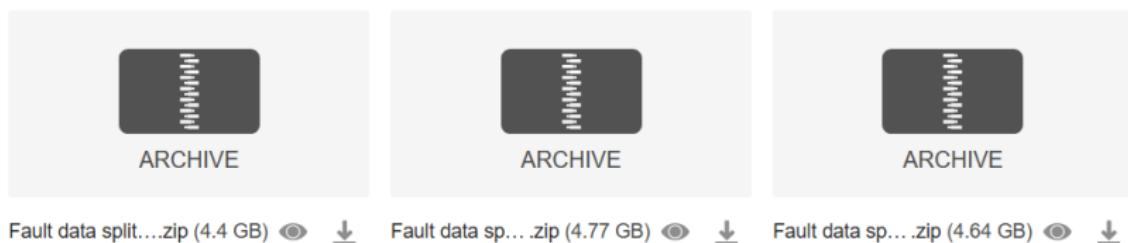


Figura: Dataset original sin descomprimir.

Introducción

- Origen de los datos: Un simulador de averías está compuesto por un motor que hace girar un eje apoyado entre dos rodamientos. Además, una polea y una correa están unidas al eje y se utilizan para accionar una caja de engranajes. Un tacómetro media las rpm del eje. Los acelerómetros piezoelectríficos se fijaron en 3 direcciones perpendiculares a las carcasa de los cojinetes y en dos direcciones en la carcasa de la caja de cambios.



Figura: Simulador de fallas

Introducción

- ▶ Descripción de la recopilación de datos: Para recopilar datos para un escenario de fallo, se reemplazó un componente nominal en el simulador de fallas de la máquina con un componente defectuoso especificado. En el caso de escenarios de dos fallos, se reemplazaron dos componentes nominales por componentes defectuosos específicos conocidos. En el caso del escenario nominal, no se sustituyó ningún componente. El motor se encendió y se ajustó a uno de los tres valores de rpm para cada escenario. Después de alcanzar una velocidad de funcionamiento constante, la recopilación de datos se realizó desde sensores durante 10 segundos. Después de la recopilación de datos, el simulador de la máquina se apagó y los componentes defectuosos se reemplazaron por los componentes nominales o por el siguiente componente defectuoso que se probaría.

Introducción



Figura: Simulador de fallas

Planteamiento del Problema

- ▶ Descripción del problema: Detección temprana de fallos en maquinaria industrial.
- ▶ Pregunta clave: ¿Cómo podemos aprovechar el deep learning para detectar fallos con precisión en maquinaria basada en datos de vibración?



Figura: Detección temprana de fallas en máquinas rotativas mediante deep learning en un entorno industrial.

Descripción de los Datos

- ▶ Visualización de muestra de datos: cada uno de los archivos contaba con 64,000 datos tomados por un acelerómetro en un periodo de 10 segundos.

# Muestra las primeras filas data.head()																		
(ipython-input-3-84613e9ff014):4: DtypeWarning: Columns (0,1,2,3,4,5,6,7,8,9,10,11,12,13,14,15,16,17) have mixed types. Specify dtype option on import or set low_memory=False. data = pd.read_csv('content/drive/MyDrive/Fault data/Fault data split 25/no Fault (real 1).csv')																		
	Timestamp	7/18/22 14:09	Timestamp.1	7/18/22 14:09.1	Timestamp.2	7/18/22 14:09.2	Timestamp.3	7/18/22 14:09.3	Timestamp.4	7/18/22 14:09.4	Timestamp.5	7/18/22 14:09.5	Timestamp.6	7/18/22 14:09.6	Timestamp.7	7/18/22 14:09.7	Timestamp.8	7/18/22 14:09.8
0	Interval	0.00015625	Interval	0.00015625	Interval	0.00015625	Interval	0.00015625	Interval	0.00015625	Interval	0.00015625	Interval	0.00015625	Interval	0.00015625	Interval	
1	Channel name	Tachometer	Channel name	Motor	Channel name	Bearing 1 Z	Channel name	Bearing 1 Y	Channel name	Bearing 1 X	Channel name	Bearing 2 Z	Channel name	Bearing 2 Y	Channel name	Bearing 2 X	Channel name	Gearbox
2	Unit	V	Unit	g	Unit													
3	0	-0.862248268	0	-0.006744791	0	0.001761446	0	-0.004679333	0	-0.013791326	0	-0.000338325	0	-0.022431522	0	0.011058366	0	0.008016799
4	0.00015625	-0.862027298	0.00015625	-0.0043829	0.00015625	0.003282426	0.00015625	-0.004456994	0.00015625	-0.01916785	0.00015625	0.00092303	0.00015625	-0.020547991	0.00015625	0.013693532	0.00015625	0.009252003

Figura: Visualización de los primeros datos de uno de los archivos originales.

Preprocesamiento de Datos

- ▶ Pasos realizados para el preprocesamiento de datos:
 - ▶ Eliminación de las tres primeras filas.
 - ▶ Eliminación de columnas innecesarias (3, 5, 7, 9, 11, 13, 15, 17).
 - ▶ Renombrar columnas a: Tiempo, Tacómetro, Motor, Rodamiento 1 Z, Rodamiento 1 Y, Rodamiento 1 X, Rodamiento 2 Z, Rodamiento 2 Y, Rodamiento 2 X, Caja de engranajes.
- ▶ Herramientas utilizadas para el preprocesamiento.

df.head()										
	Time	Tachometer	Motor	Bearing 1 Z	Bearing 1 Y	Bearing 1 X	Bearing 2 Z	Bearing 2 Y	Bearing 2 X	Gearbox
0	0	0.001317648384999992	0.0082865870495040399	0.021089014198019798	-0.0023447754257425752	-0.015215976540000016	0.0047812919603960335	-0.024906489900990106	0.018778299811881191	0.011759840489999979
1	0.00015625	0.001074206809999997	0.0077082636633263177	0.014745417039603959	-0.0021234365336633667	-0.014242934330000013	0.0057829561386183853	-0.029819486198019808	0.0129883467277277	0.009776628759999857
2	0.0003125	0.001077952064999994	0.003917861363366224	0.01259378794059404	-0.0067174383366336646	-0.006818035100000015	-0.002675541366336631	-0.0259038665091092	0.01833299574257428	0.011310675569999984
3	0.00046875	0.001145366624999996	0.0070648247821782084	0.01599691999999997	-0.0040864292970297037	-0.012719220930000012	0.0019246941188118746	-0.02852659243564373	0.01859274712871286	0.012583309509999988
4	0.000625	0.001182819204999998	0.0077682636633263177	0.013187340544554452	-0.0070138900594059412	-0.013228963610000012	0.0058200548118811854	-0.026141252891089118	0.006567448297029703	0.006182709399999986

Figura: Visualización de primeros datos de uno de los archivos optimizados.

Visualización de datos

► Series temporales

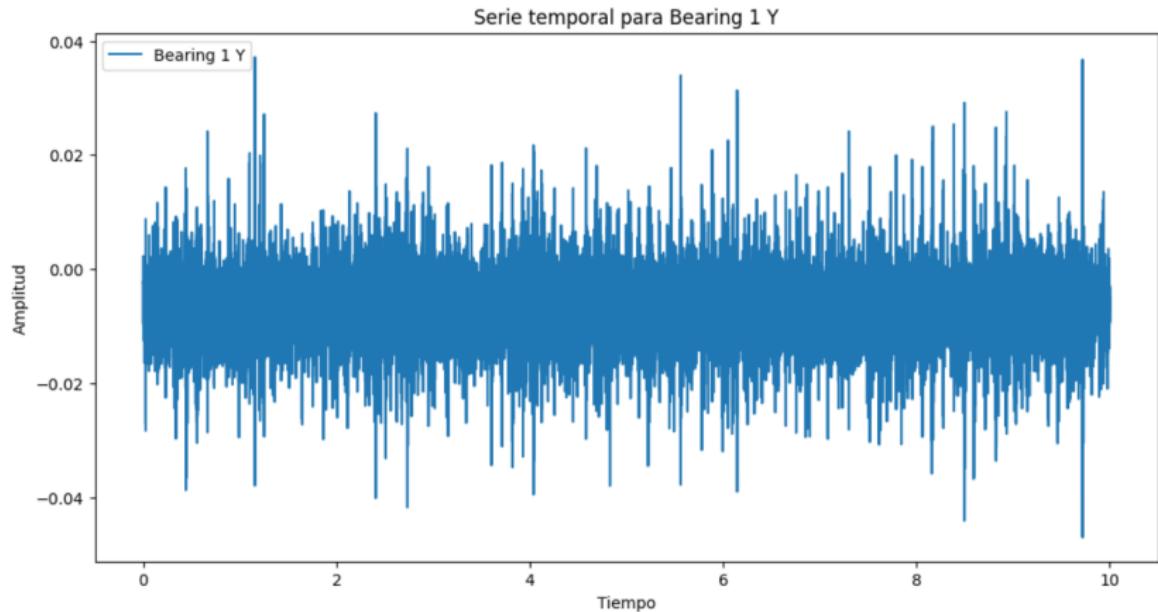


Figura: Serie temporal para cojinete en una de las direcciones

Visualización de datos

- ▶ Dominio de frecuencias del cojinete en uno de los escenarios sin falla.

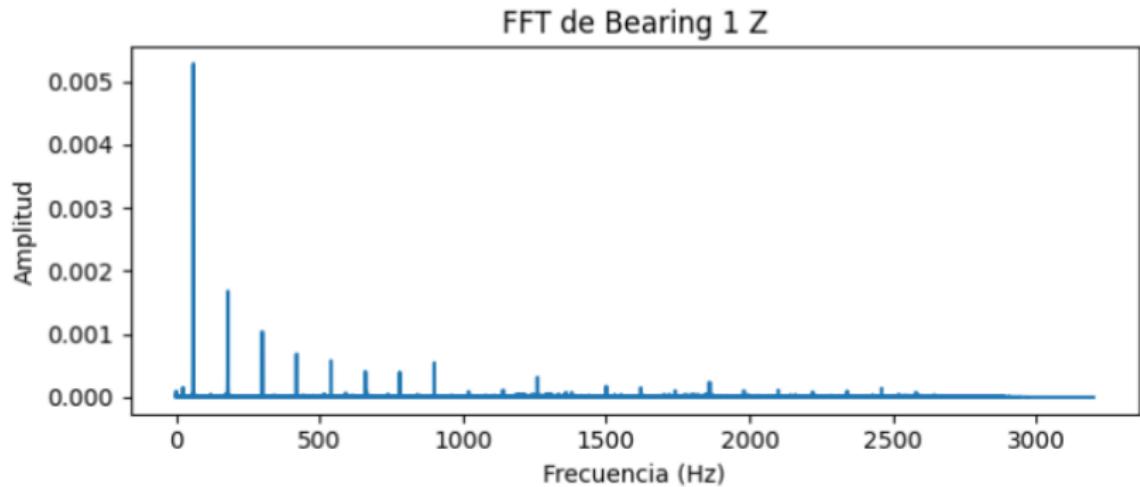


Figura: Dominio de frecuencia mediante FFT

Visualización de datos

- ▶ Dominio de frecuencias del cojinete en uno de los escenarios con falla.

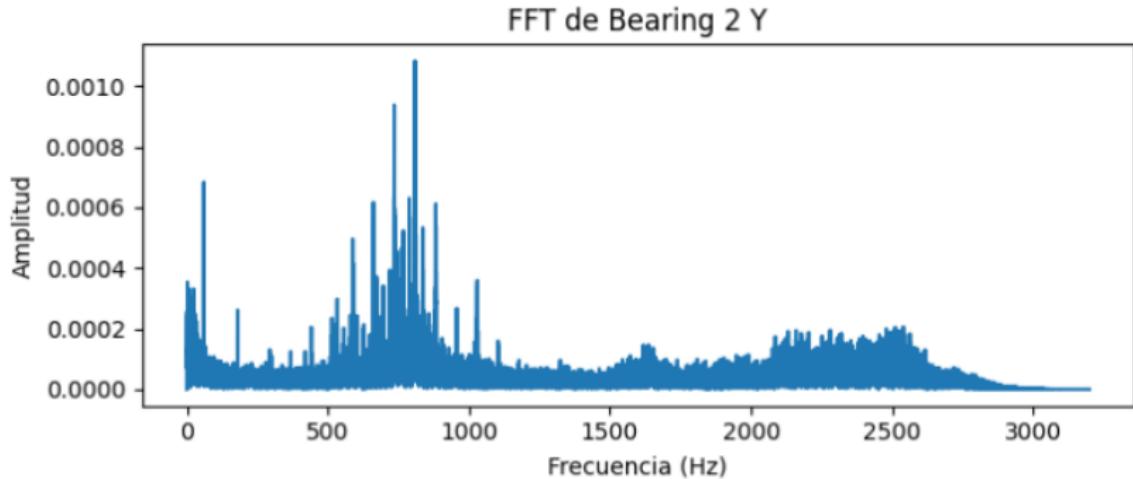


Figura: Dominio de frecuencia mediante FFT

Extracción de Características

- ▶ Explicación del proceso de extracción de características.
- ▶ Características clave consideradas para el análisis:
Características en el dominio del tiempo, características en el dominio de la frecuencia.
- ▶ Importancia de cada característica en la detección de fallos.

df_features.head()																		column	file	late	
	mean	std	max	min	ptp	kurtosis	skewness	rms	crest_factor	fft_mean	...	harmonic_3_amplitude	harmonic_4_amplitude	harmonic_5_amplitude	envelope_mean	envelope_std	wavelet_mean	wavelet_std	column	file	late
0	0.000518	0.000420	0.002520	-0.000791	0.003211	0.491388	0.280136	0.000674	3.737075	3.382358e-07	...	0.000518	0.000518	0.000318	0.000724	0.000342	0.358662e-09	0.000190	Tachometer	(Fault) Bearing (1) Ball & Bearing (2) Combina...	1
1	-0.001712	0.004595	0.012692	-0.019407	0.032099	-0.013154	0.008711	0.004904	3.957529	4.663926e-06	...	0.000918	0.000006	0.000243	0.006217	0.002552	8.508267e-08	0.001379	Motor	(Fault) Bearing (1) Ball & Bearing (2) Combina...	1
2	0.006132	0.005580	0.062412	-0.041879	0.102091	1.048149	-0.029990	0.008290	7.287653	1.059404e-05	...	0.006132	0.006132	0.000982	0.004389	9.672096e-08	0.003565	Bearing Y Z	(Fault) Bearing (1) Ball & Bearing (2) Combina...	1	
3	0.006815	0.004825	0.037157	-0.047035	0.084192	4.766746	-0.286269	0.008181	5.749631	1.212229e-05	...	0.006815	0.006815	0.000383	0.004139	4.416659e-08	0.004398	Bearing Y X	(Fault) Bearing (1) Ball & Bearing (2) Combina...	1	
4	-0.015174	0.006319	0.066778	-0.076701	0.157479	6.155259	-0.001890	0.016434	4.667125	1.533595e-05	...	0.015174	0.015174	0.016581	0.009916	2.849377e-10	0.007329	Bearing 1 X	(Fault) Bearing (1) Ball & Bearing (2) Combina...	1	

Figura: Extracción de features de los 975 archivos para un régimen de 25 RPM.

Modelo Preliminar CNN

- ▶ Resumen de las capas y parámetros del modelo.

Layer (type)	Output Shape	Param #
=====		
conv1d_4 (Conv1D)	(None, 98, 64)	256
max_pooling1d_4 (MaxPooling1D)	(None, 49, 64)	0
conv1d_5 (Conv1D)	(None, 47, 128)	24704
max_pooling1d_5 (MaxPooling1D)	(None, 23, 128)	0
flatten_2 (Flatten)	(None, 2944)	0
dense_4 (Dense)	(None, 100)	294500
dense_5 (Dense)	(None, 1)	101
=====		
Total params: 319561 (1.22 MB)		
Trainable params: 319561 (1.22 MB)		
Non-trainable params: 0 (0.00 Byte)		

Figura: Estructura del modelo en formato de texto

Modelo Preliminar CNN

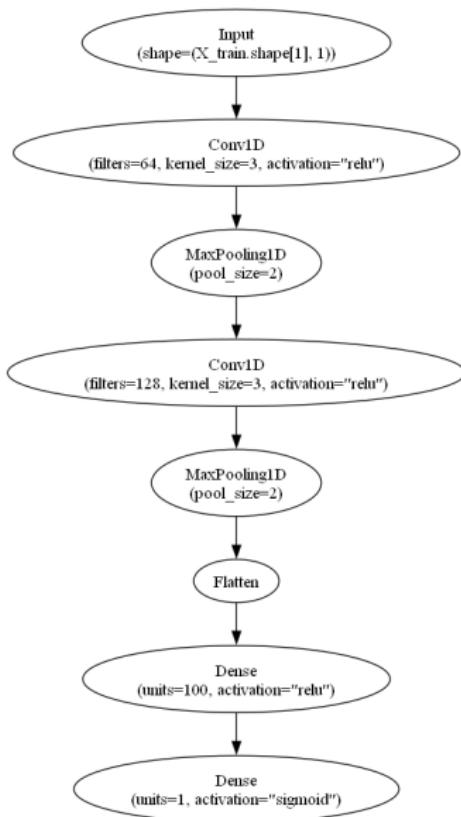


Figura: Estructura del modelo

Entrenamiento y Evaluación del Modelo

- ▶ Primeramente el modelo fue entrenado con todos los escenarios (38 de fallas y 1 solo de no falla).
- ▶ Resultados preliminares y métricas de rendimiento.

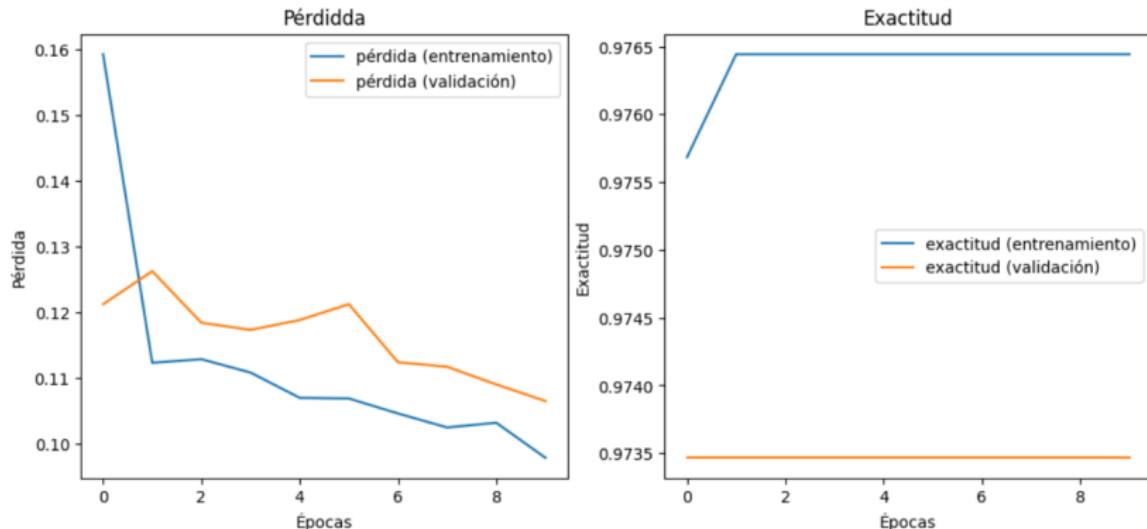


Figura: Pérdida y exactitud vs. cantidad de épocas, para los conjuntos de validación y entrenamiento.

Entrenamiento y Evaluación del Modelo

- ▶ Resultados preliminares y métricas de rendimiento.

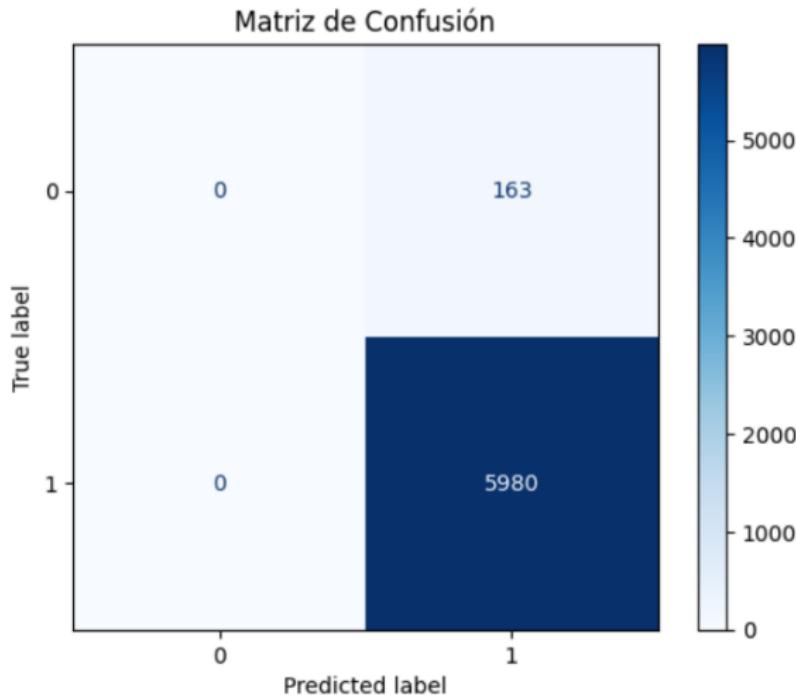


Figura: Matriz de confusión con todos los escenarios.

Entrenamiento y Evaluación del Modelo

- ▶ El modelo estaba claramente sobreajustado, por lo que balanceamos la cantidad de datos comparando 1 falla vs. 1 no falla.

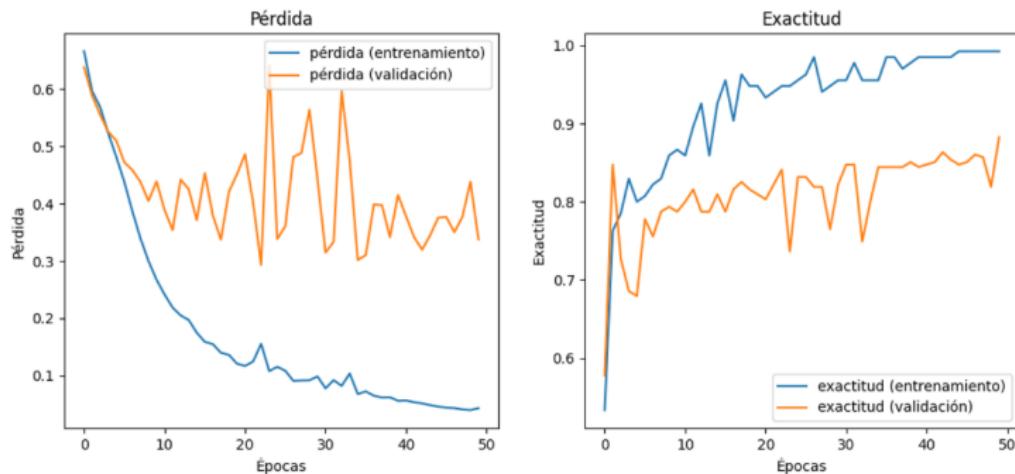


Figura: Pérdida y exactitud vs. cantidad de épocas, para los conjuntos de validación y entrenamiento 25 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ Resultados preliminares y métricas de rendimiento.

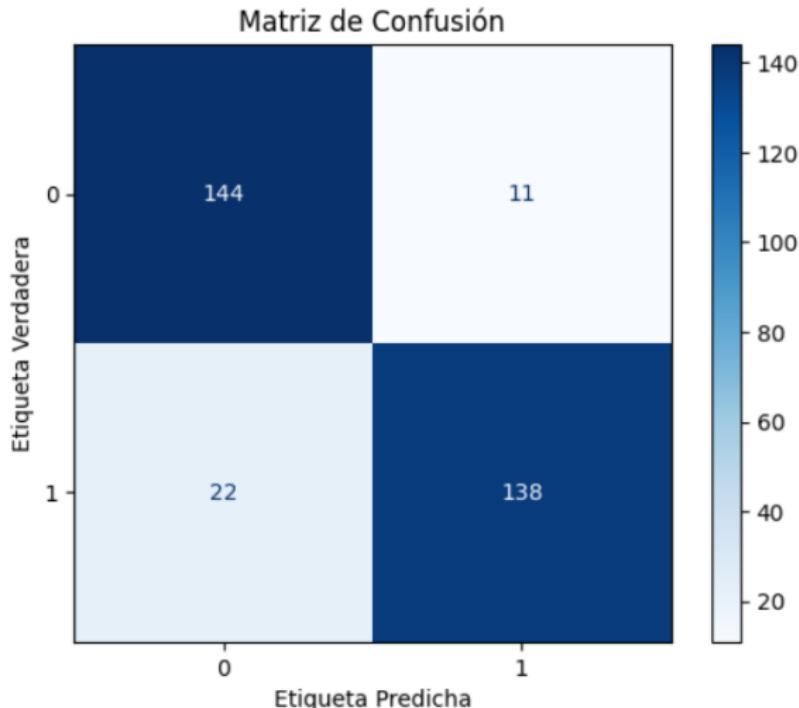


Figura: Matriz de confusión 25 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ Resultados preliminares y métricas de rendimiento.

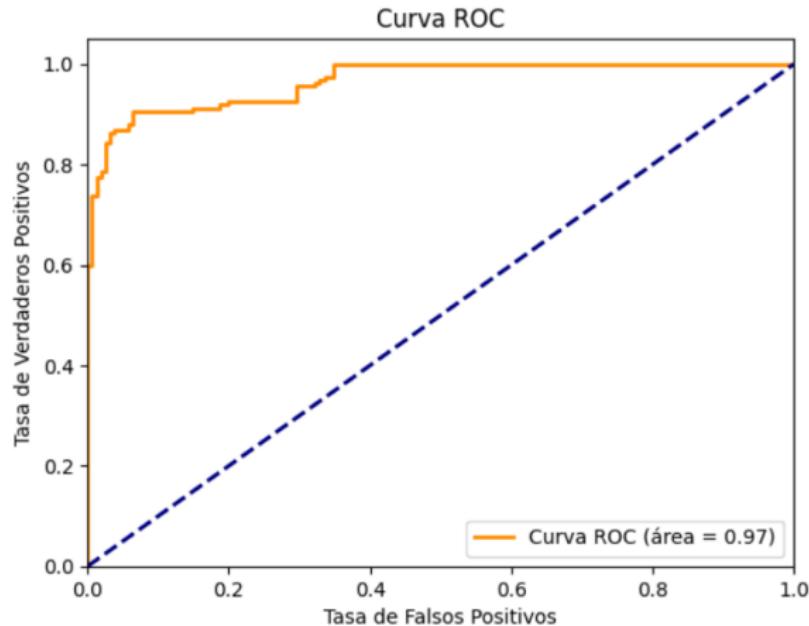


Figura: Curva ROC 25 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ Resultados preliminares y métricas de rendimiento.

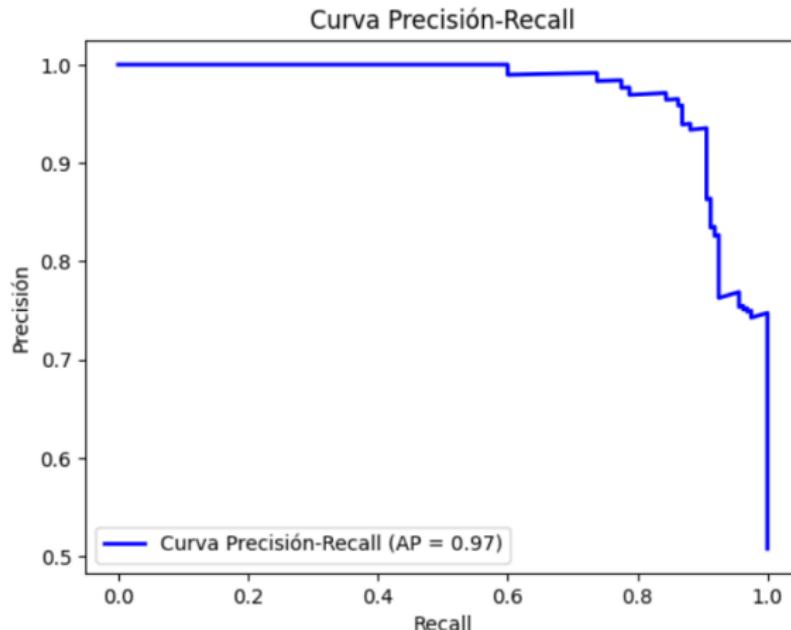


Figura: Curva de Precisión - Recall 25 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ El modelo estaba claramente sobreajustado, por lo que balanceamos la cantidad de datos comparando 1 falla vs. 1 no falla.

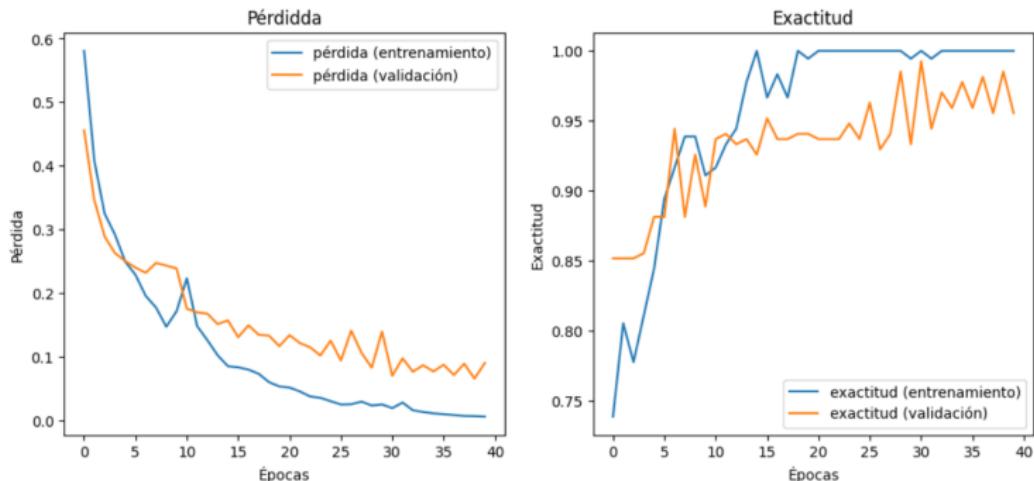


Figura: Pérdida y exactitud vs. cantidad de épocas, para los conjuntos de validación y entrenamiento 50 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ Resultados preliminares y métricas de rendimiento.

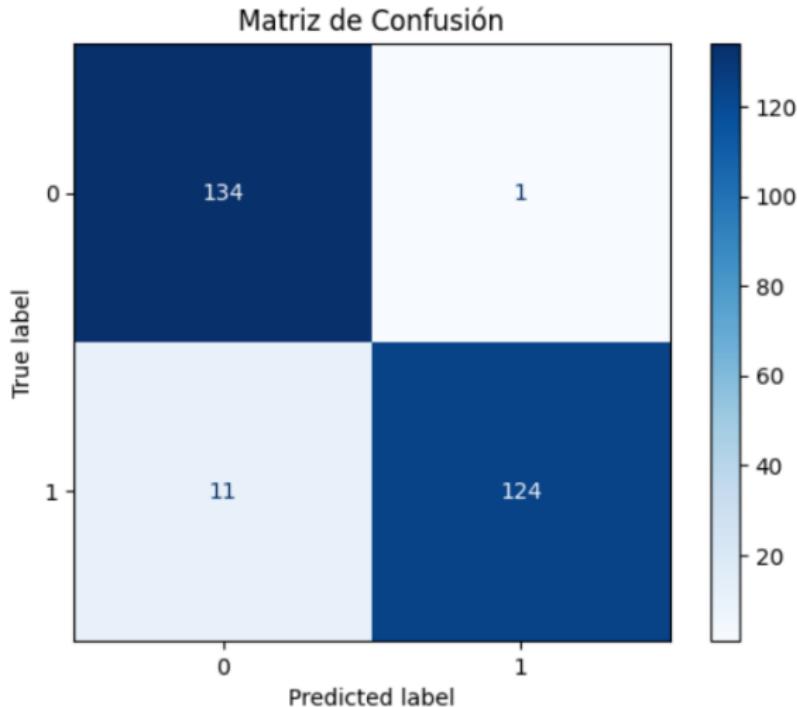


Figura: Matriz de confusión 50 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ Resultados preliminares y métricas de rendimiento.

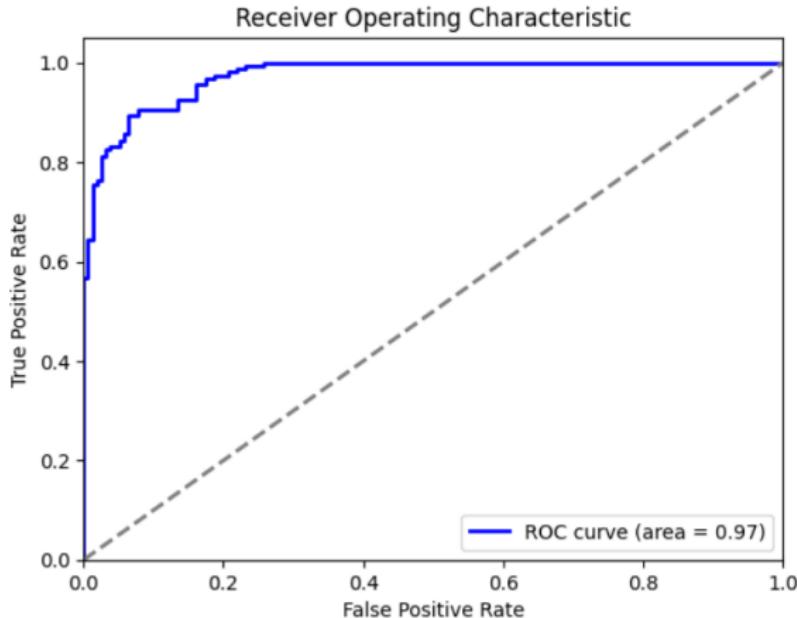


Figura: Curva ROC 50 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ Resultados preliminares y métricas de rendimiento.

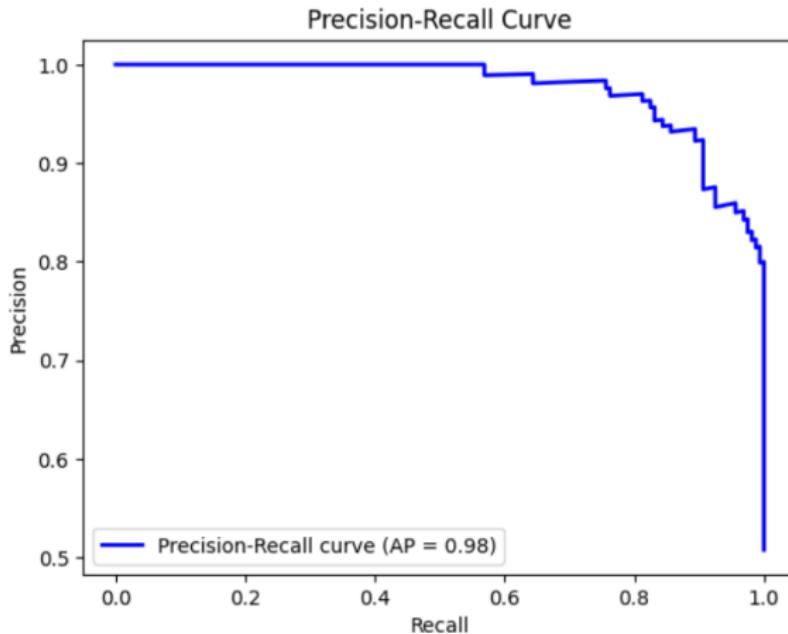


Figura: Curva de Precisión - Recall 50 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ El modelo estaba claramente sobreajustado, por lo que balanceamos la cantidad de datos comparando 1 falla vs. 1 no falla.

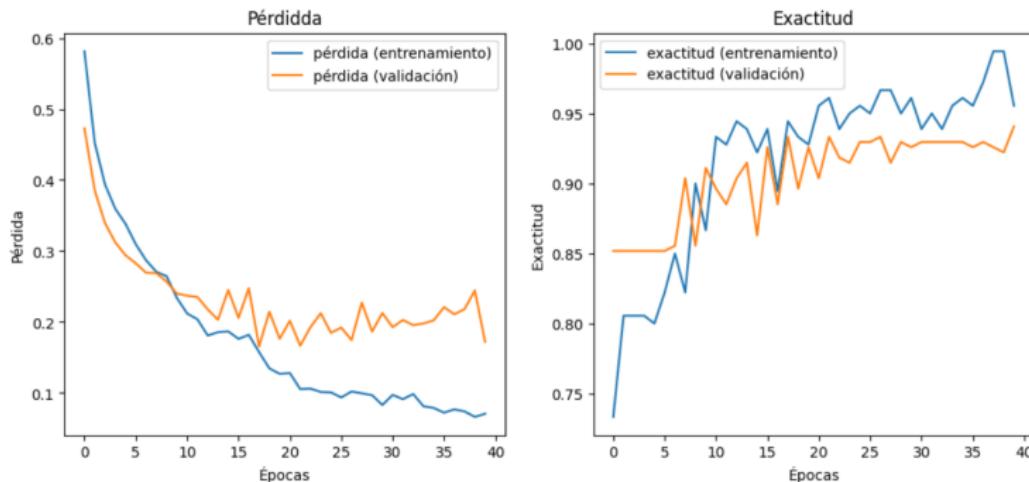


Figura: Pérdida y exactitud vs. cantidad de épocas, para los conjuntos de validación y entrenamiento 75 RPM.

Entrenamiento y Evaluación del Modelo

- ▶ Resultados preliminares y métricas de rendimiento.

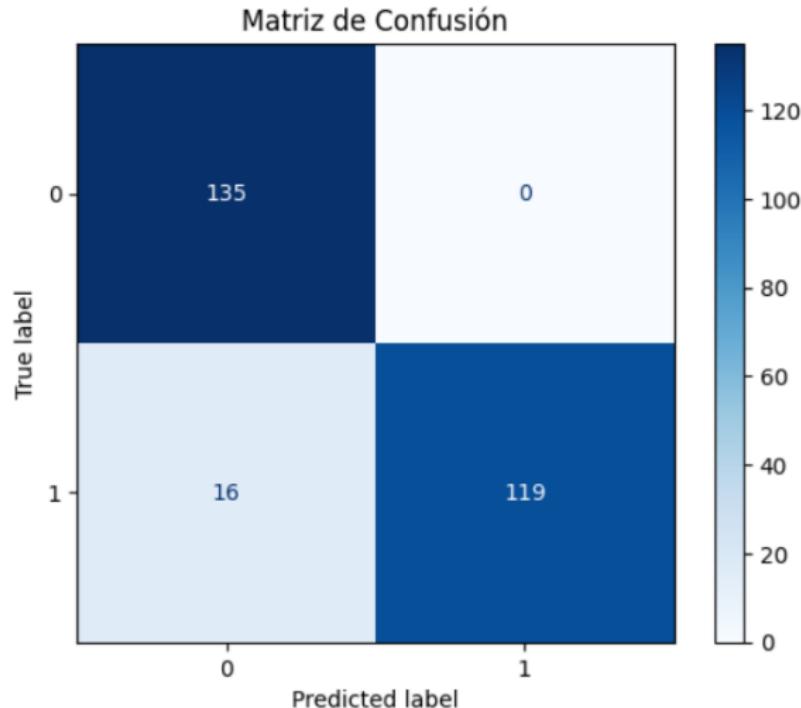


Figura: Matriz de confusión 75 RPM.

Mejoras para el pre-procesamiento y feature engineering

- ▶ Corrección en la extracción de características.
- ▶ Visualización de los primeros 10 armónicos.
- ▶ Verificación de la influencia de los armónicos en las fallas.

Pre-procesamiento

- ▶ Se extrajeron los datos de los primeros diez armónicos para cada columna de interés (Tacómetro (1), Cojinetes (6), Motor (1), Caja de engranajes(1)).
- ▶ El objetivo fundamental es visualizar los datos más relevantes previos a la selección del modelo.

Feature engineering - Visualización

- ▶ Primeramente ploteamos los armónicos de los escenarios de falla y no falla superpuestos.
- ▶ En este caso, para el tacómetro no se visualizan diferencias relevantes entre las fallas y no fallas.

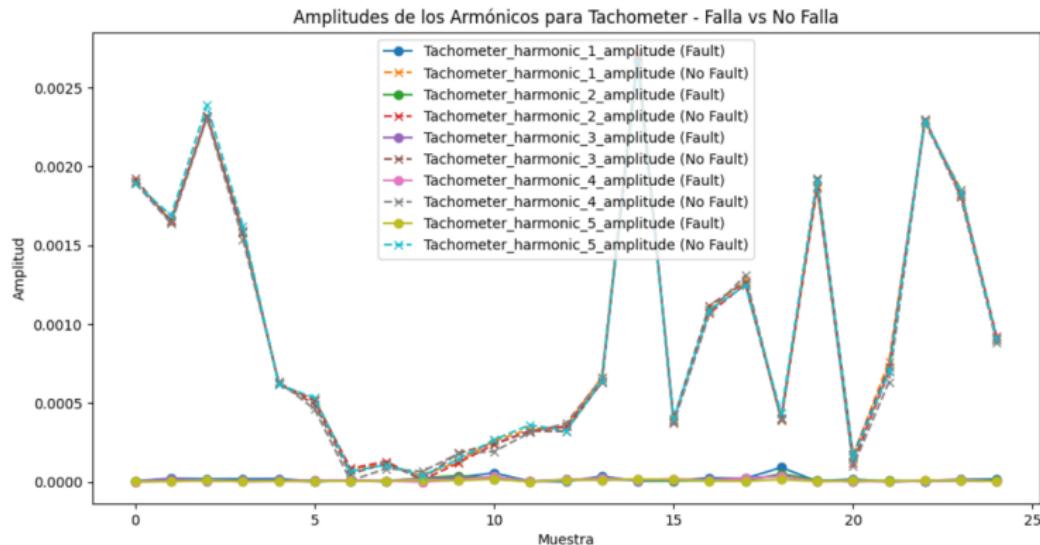


Figura: Armónicos de falla y no falla superpuestos para el tacómetro.

Feature engineering - Visualización

- ▶ Primeramente ploteamos los armónicos de los escenarios de falla y no falla superpuestos.
- ▶ Para las demás muestras, si se aprecian diferencias significativas entre los casos de falla y no falla.

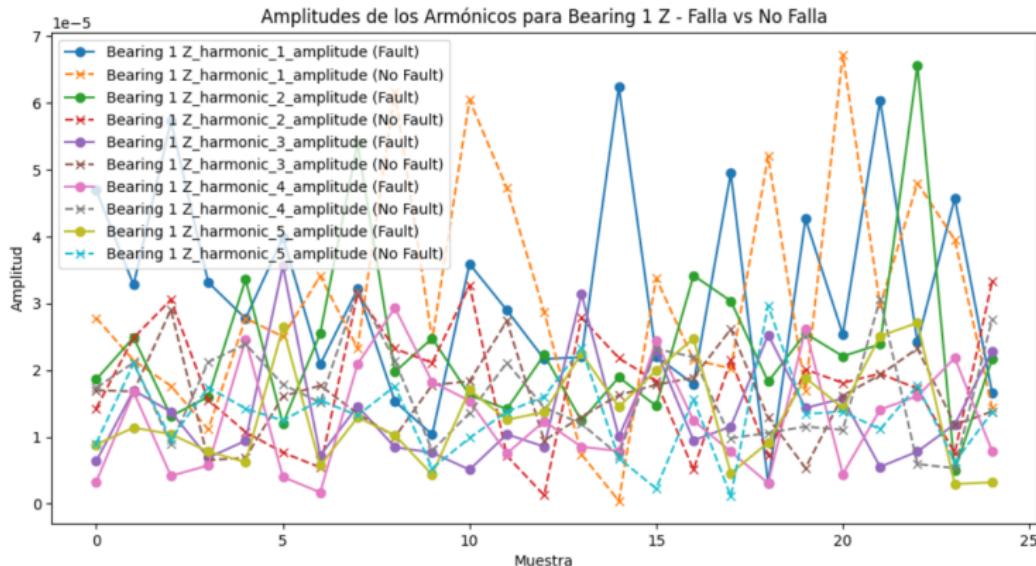


Figura: Armónicos de falla y no falla superpuestos para el cojinete 1, dirección Z.

Feature engineering - Visualización

- ▶ Para cada uno de los escenarios, hallamos las medias de los picos de los armónicos y comparamos dichas medias entre las fallas y no fallas. Click aquí para una mejor visualización.

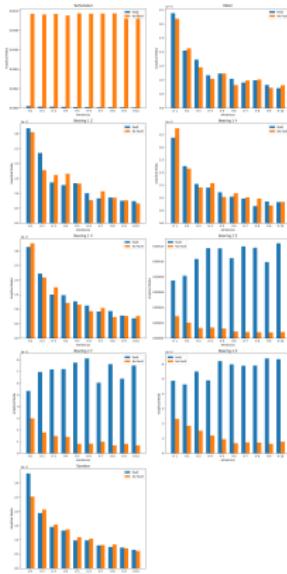


Figura: Comparativa de las medias de los primeros diez armónicos entre las fallas y no fallas.

Feature engineering - Visualización

- ▶ En el frame anterior se visualiza uno de los escenarios de falla (falla simple cojinete 1, falla combinada cojinete 2).
- ▶ Se precian diferencias importantes en las medias en las direcciones x, y y z del cojinete 2 (falla combinada).
- ▶ Se repite este proceso con los demás escenarios observándose así similitudes al gráfico mostrado anteriormente (diferencias significativas en el cojinete donde existen las fallas múltiples).

Feature engineering - Visualización

- ▶ Ahora, como los picos de los armónicos de fallas y no fallas, no necesariamente coinciden, se analiza otra alternativa.

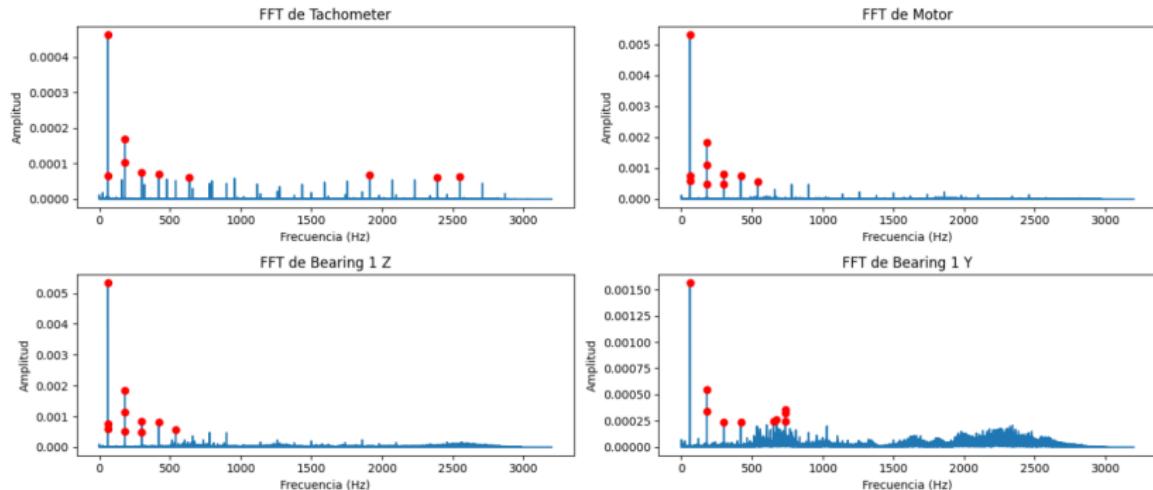


Figura: Visualización de armónicos.

Feature engineering - Visualización

- ▶ La nueva alternativa es identificar las frecuencias de los picos en los armónicos de la FFT en las fallas y comparar con el valor de la amplitud para esas mismas frecuencias en las no fallas. Click aquí para una mejor visualización.



Figura: Comparativa de las medias de los primeros diez armónicos entre las fallas y no fallas (modificado).

Feature engineering - Visualización

- ▶ Se repiten los experimentos para los demás escenarios, obteniéndose así resultados similares.
- ▶ Para estos casos modificados, se aprecia una mejora significativa a la diferenciación entre las medias de los picos de armónicos comparativos de las fallas y no fallas en algunos escenarios.
- ▶ La diferenciación se nota mejor en los cojinetes donde existen fallas múltiples.

Conclusiones y futuras mejoras

- ▶ Buscar más alternativas para la extracción de características adecuadas para la mejor selección del modelo predictivo.
- ▶ Verificar sensibilidad de las variables para una mejor visualización y selección de las características.