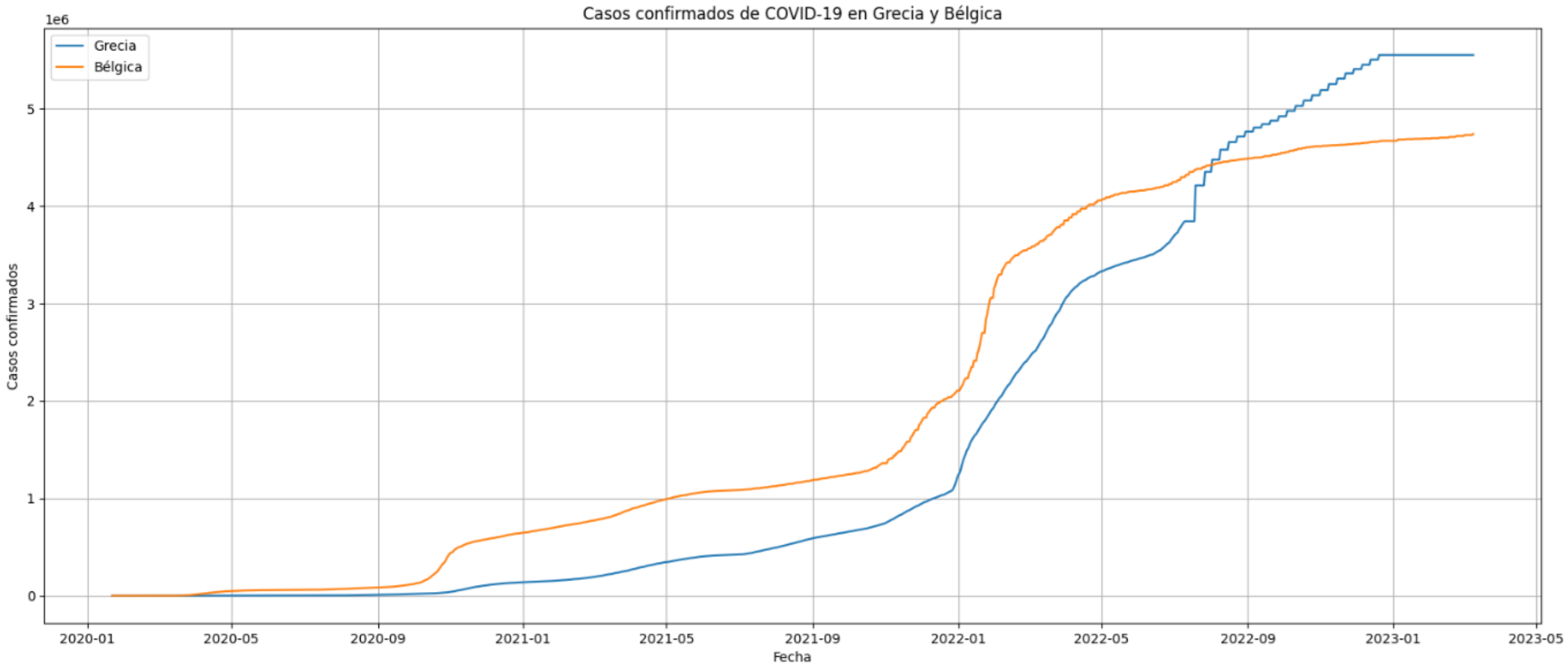


PRÁCTICA PEC3 ENTREGABLE: PREDICCIÓN DINÁMICAS DE COVID- 19 MEDIANTE REDES NEURONALES ARTIFICIALES

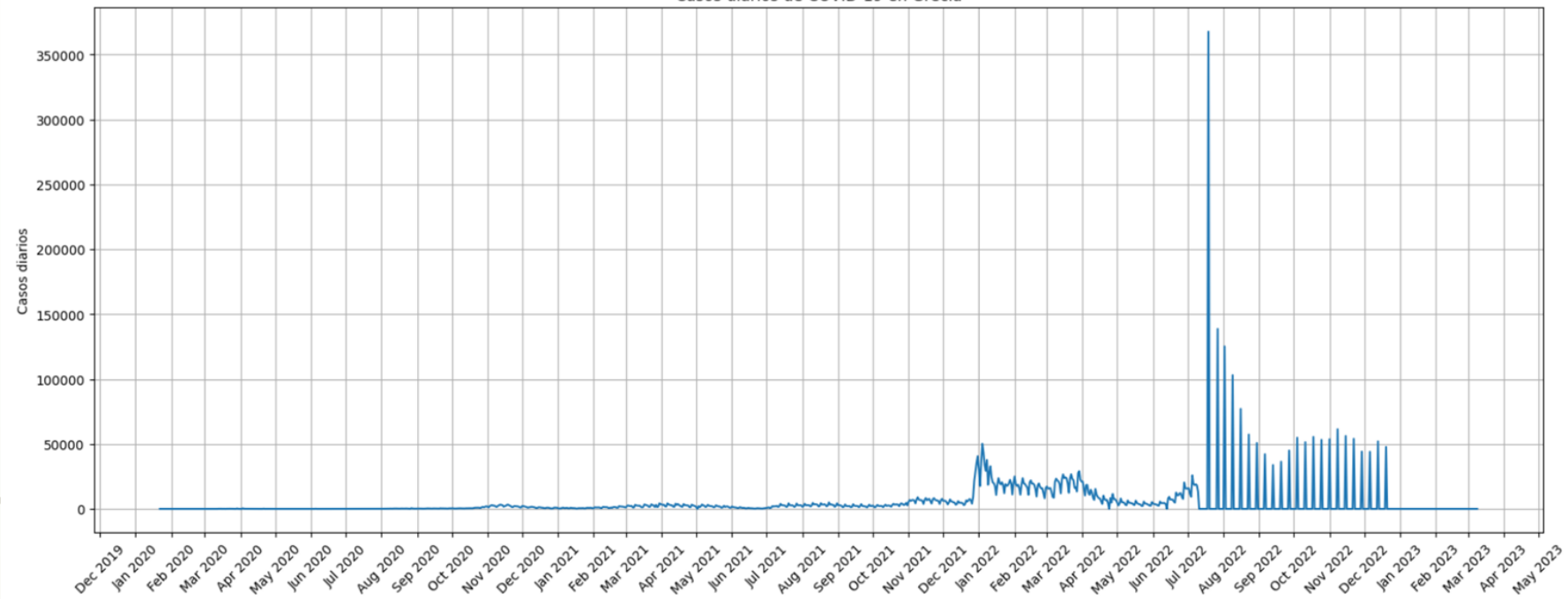


Rubén Adarve Pérez
Ángela Noelia Agreda Jaramillo
Lucía García Migueláñez
Francisco Javier Redondo García

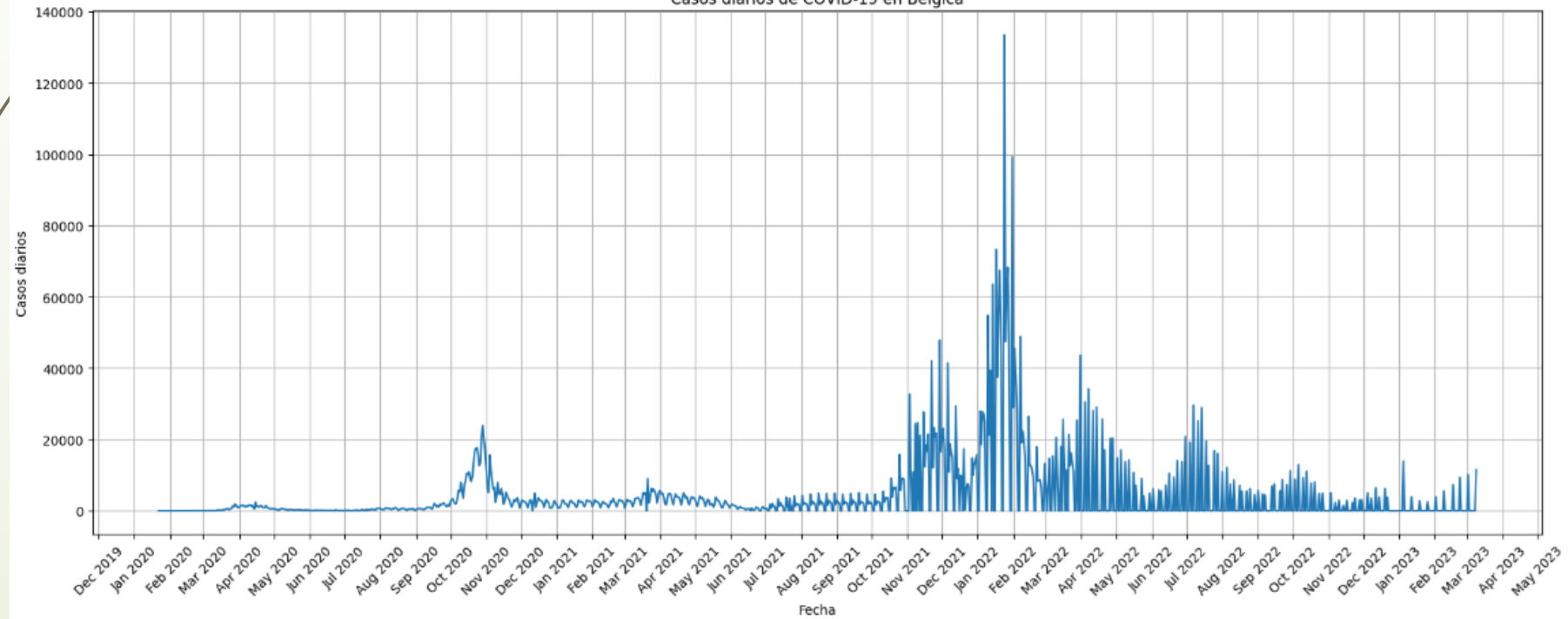
Introducción al problema



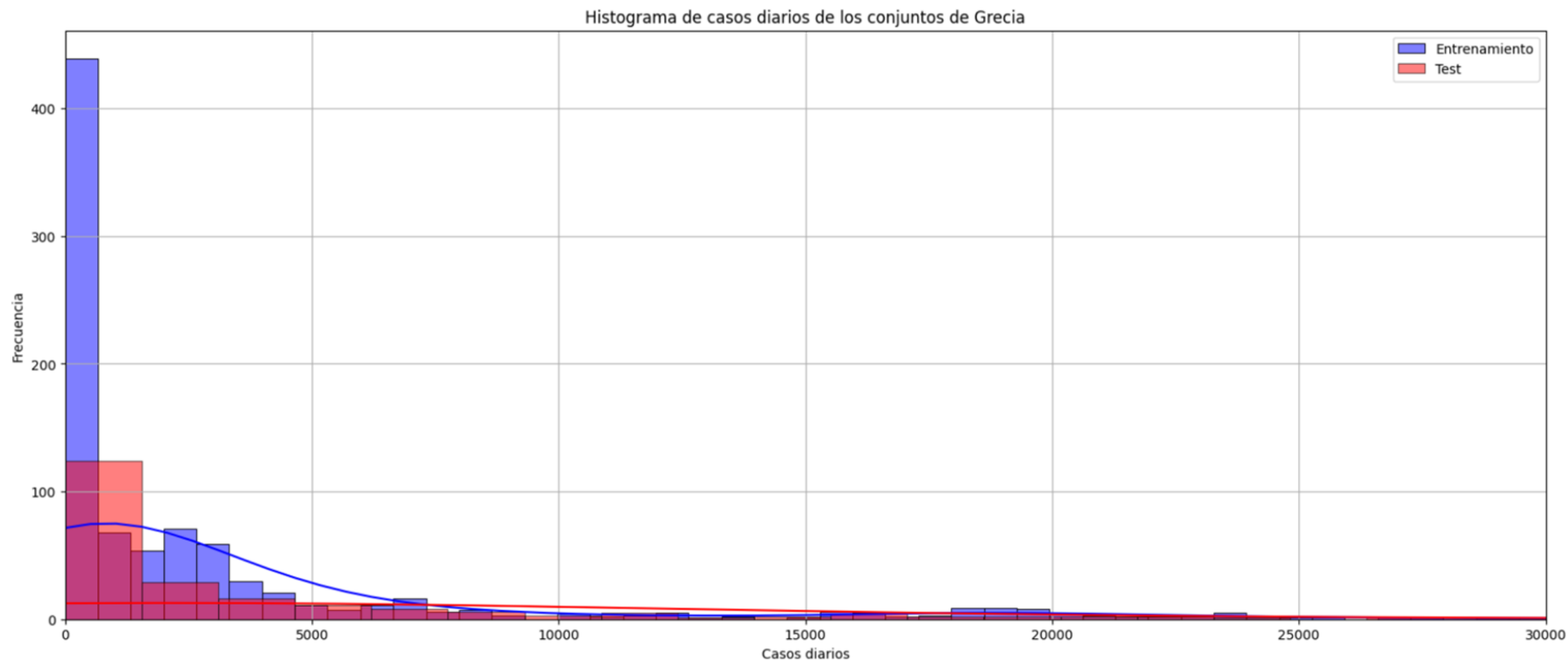
Casos diarios de COVID-19 en Grecia



Casos diarios de COVID-19 en Bélgica



1. Análisis del conjunto: entrenamiento y test



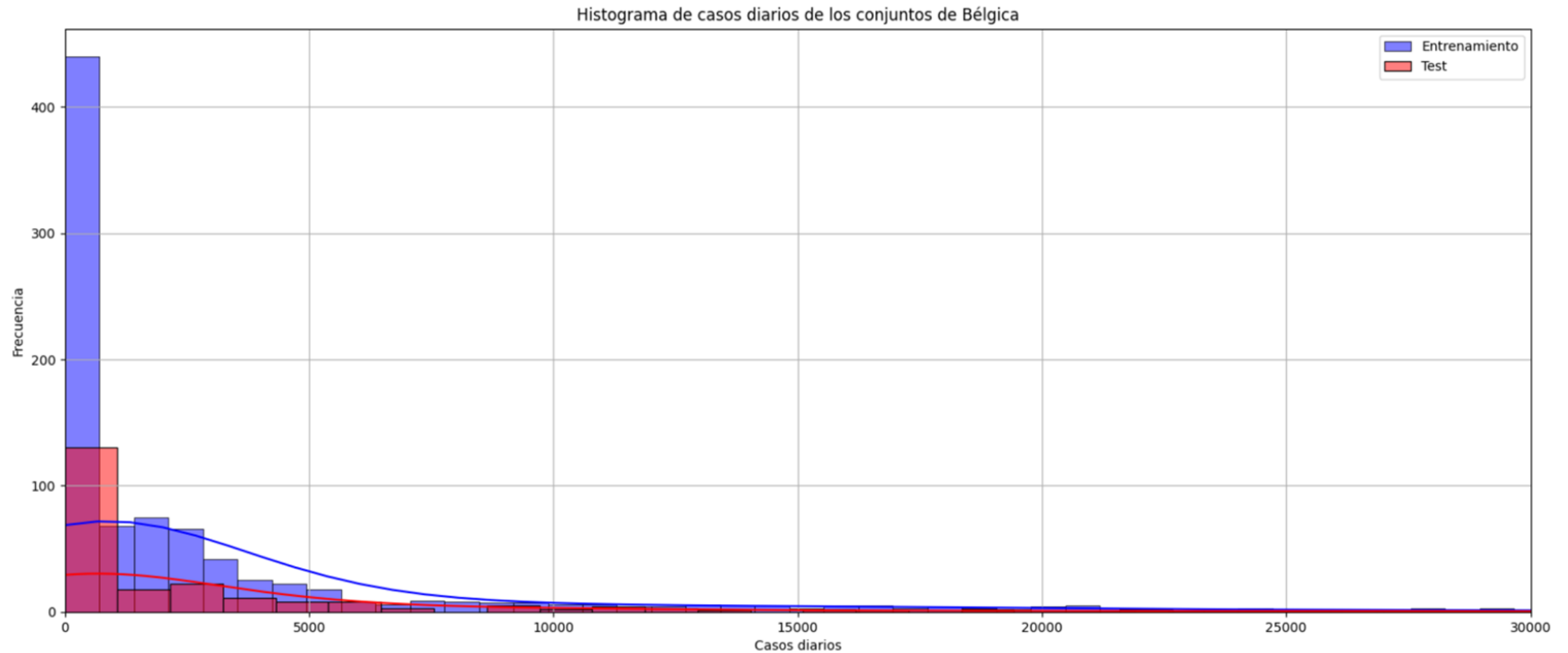
Estadísticas de los conjuntos de entrenamiento y test:

	count	mean	std	min	25%	50%	75%	max
0	913.0	4180.857612	9072.187504	0.0	2.0	833.0	3237.0	103059.0
1	229.0	7560.541485	28606.565452	0.0	3.0	1307.0	4761.0	367629.0

Sesgo del conjunto de entrenamiento: 4.200254231400659

Sesgo del conjunto de test: 9.644812534694037

1. Análisis del conjunto: entrenamiento y test



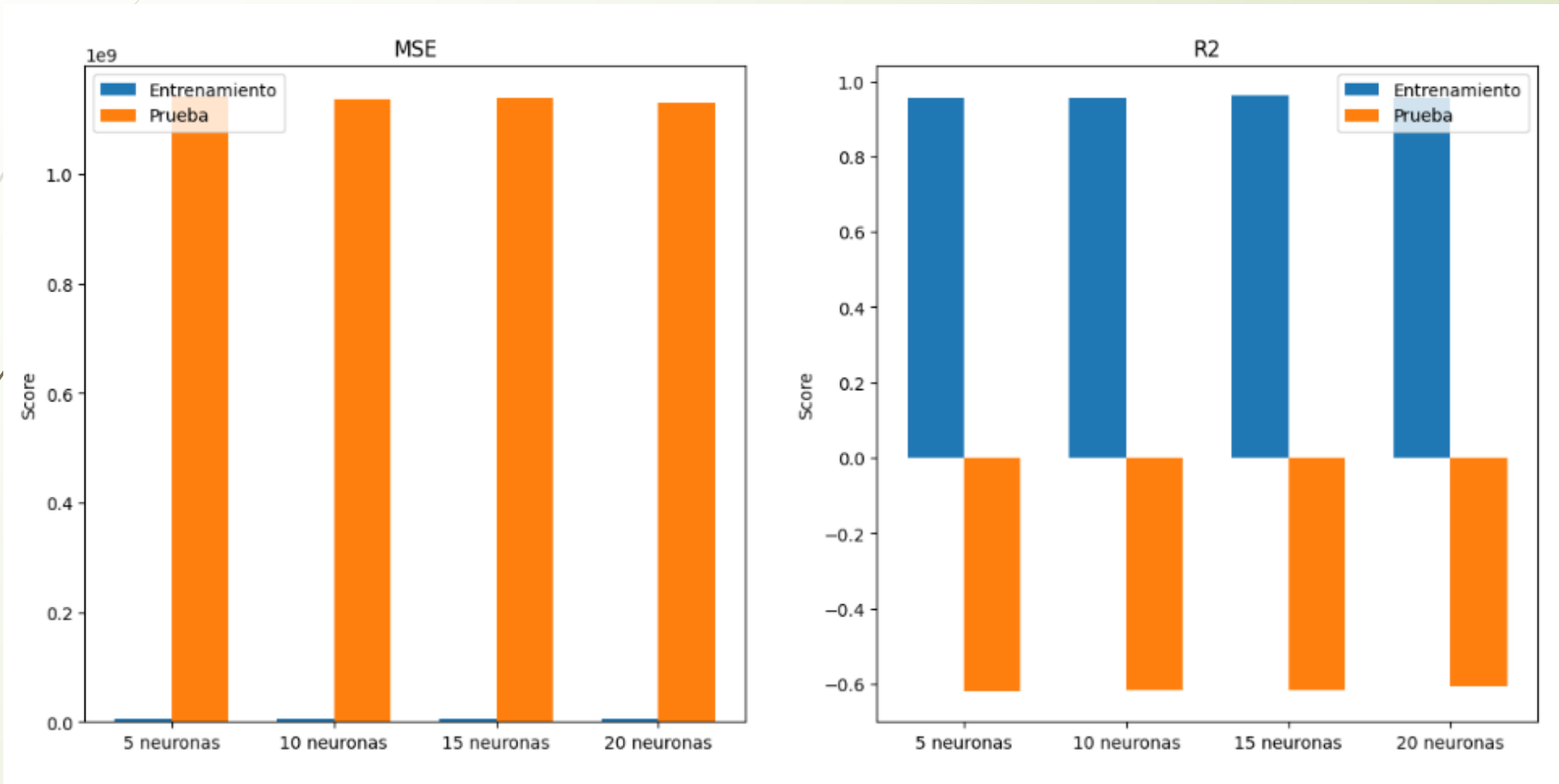
Estadísticas de los conjuntos de entrenamiento y test:

	count	mean	std	min	25%	50%	75%	max
0	913.0	4333.184009	10299.456756	0.0	0.0	857.0	3430.0	133480.0
1	229.0	3419.947598	6921.965001	0.0	0.0	528.0	3374.0	47493.0

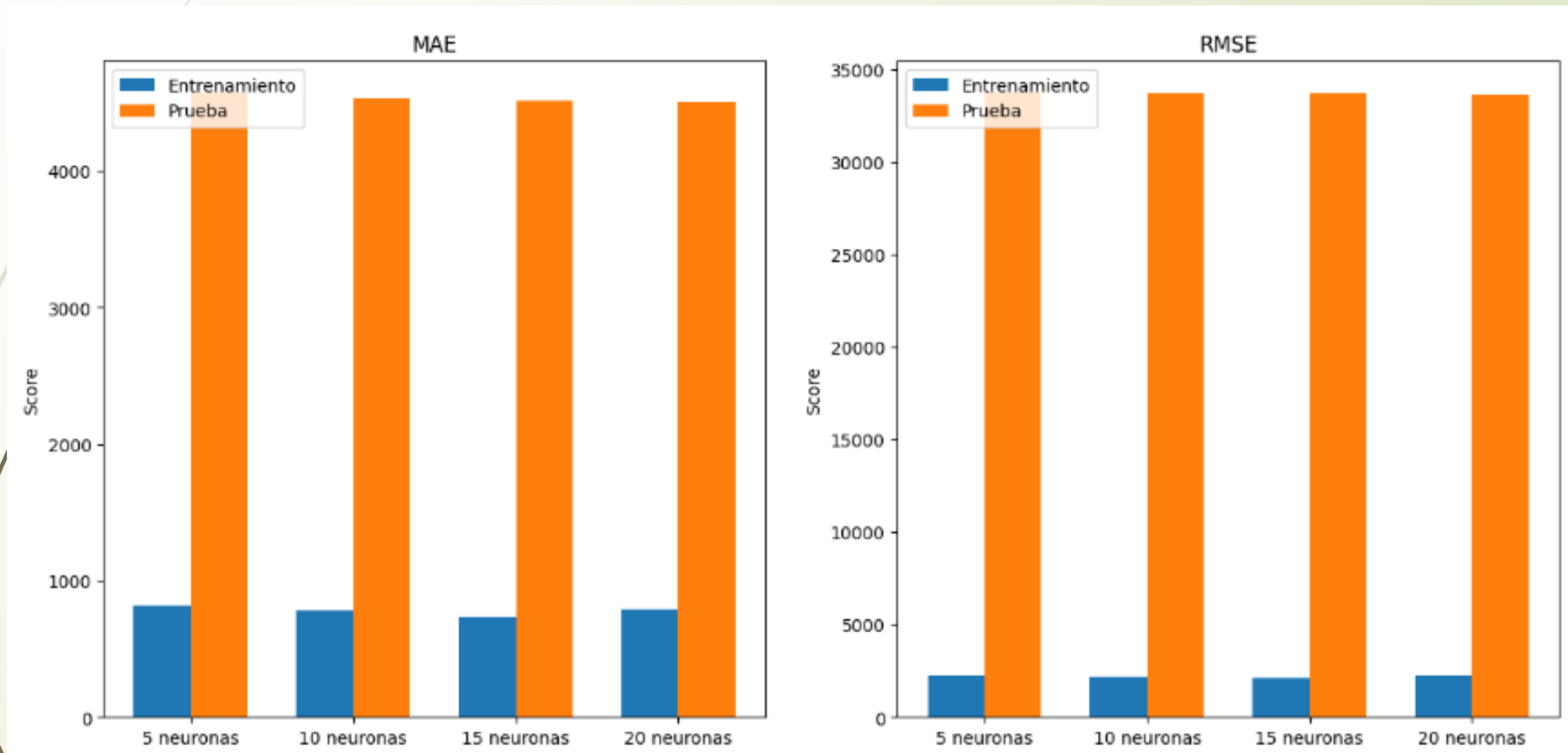
Sesgo del conjunto de entrenamiento: 5.523756848867932

Sesgo del conjunto de test: 3.4897285277023324

2. Red alimentada: 5,10,15 y 20 neuronas



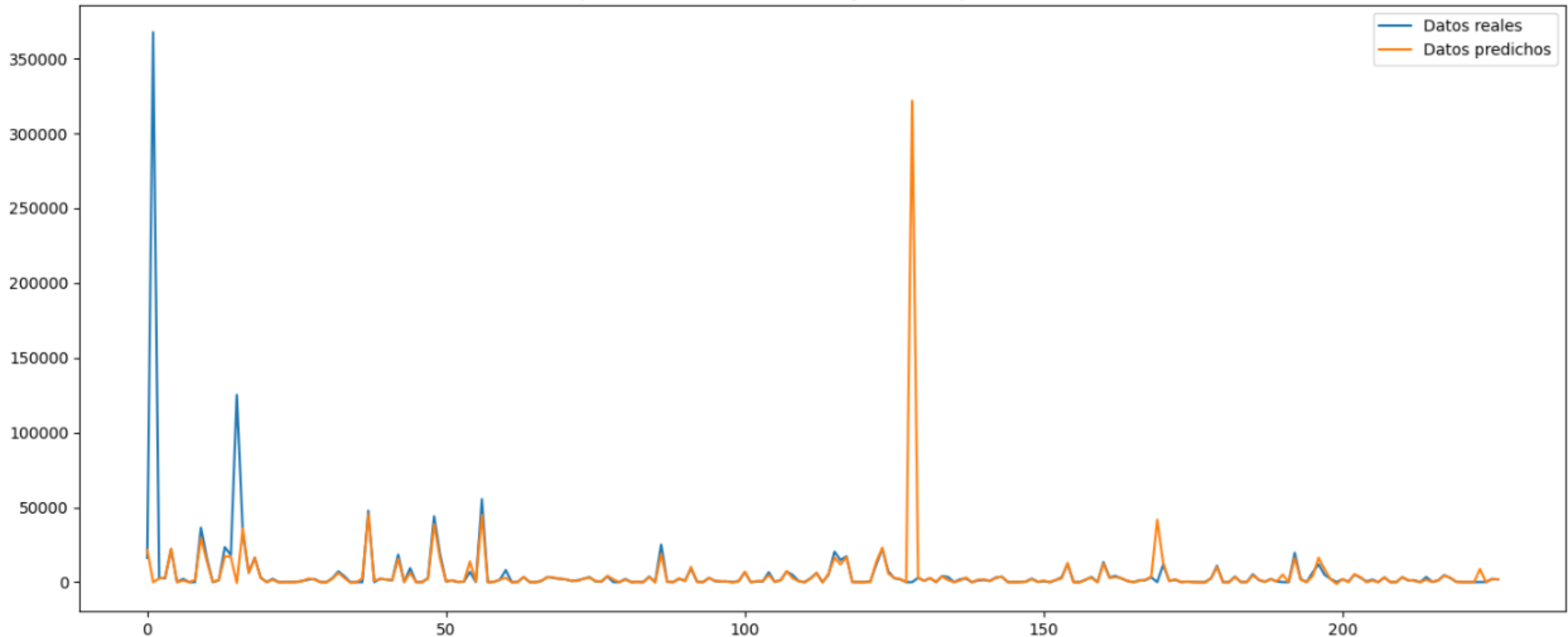
2. Red alimentada: 5,10,15 y 20 neuronas



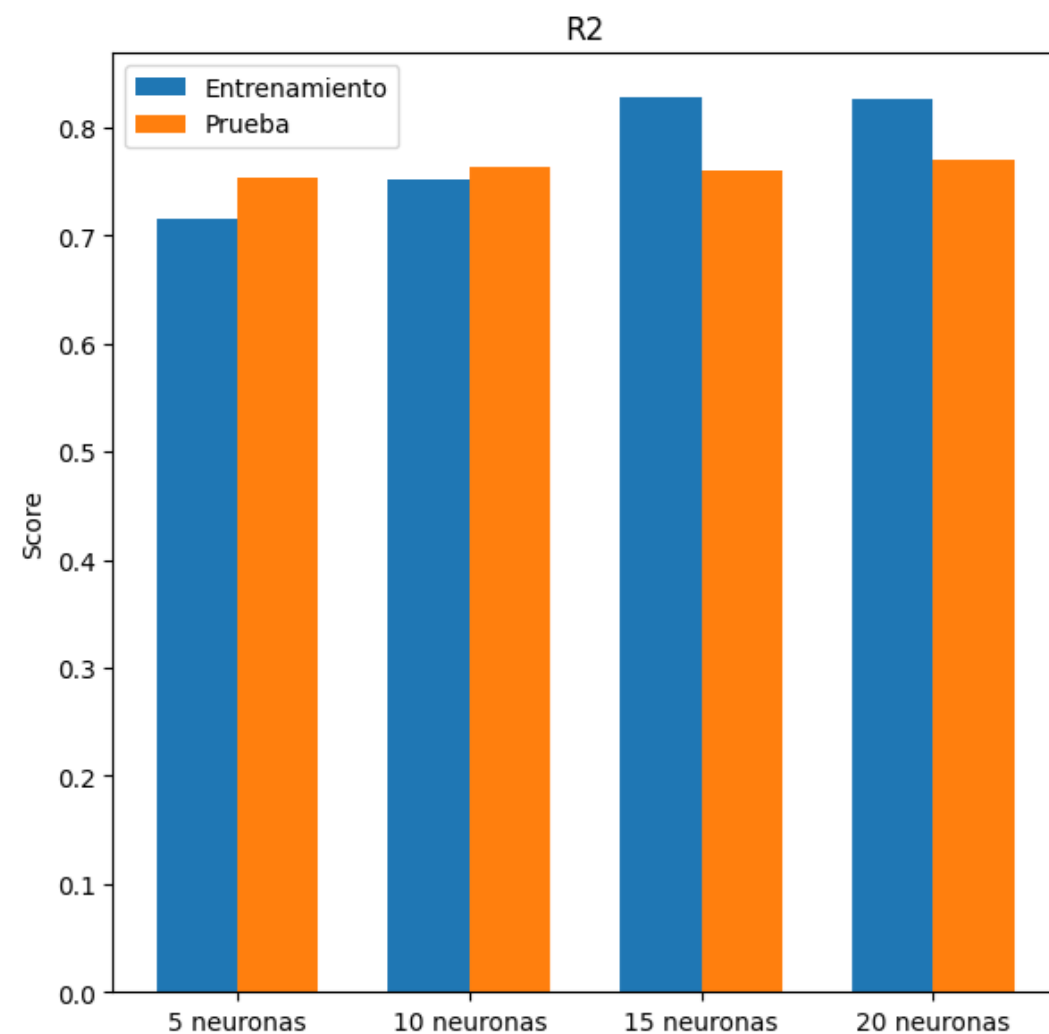
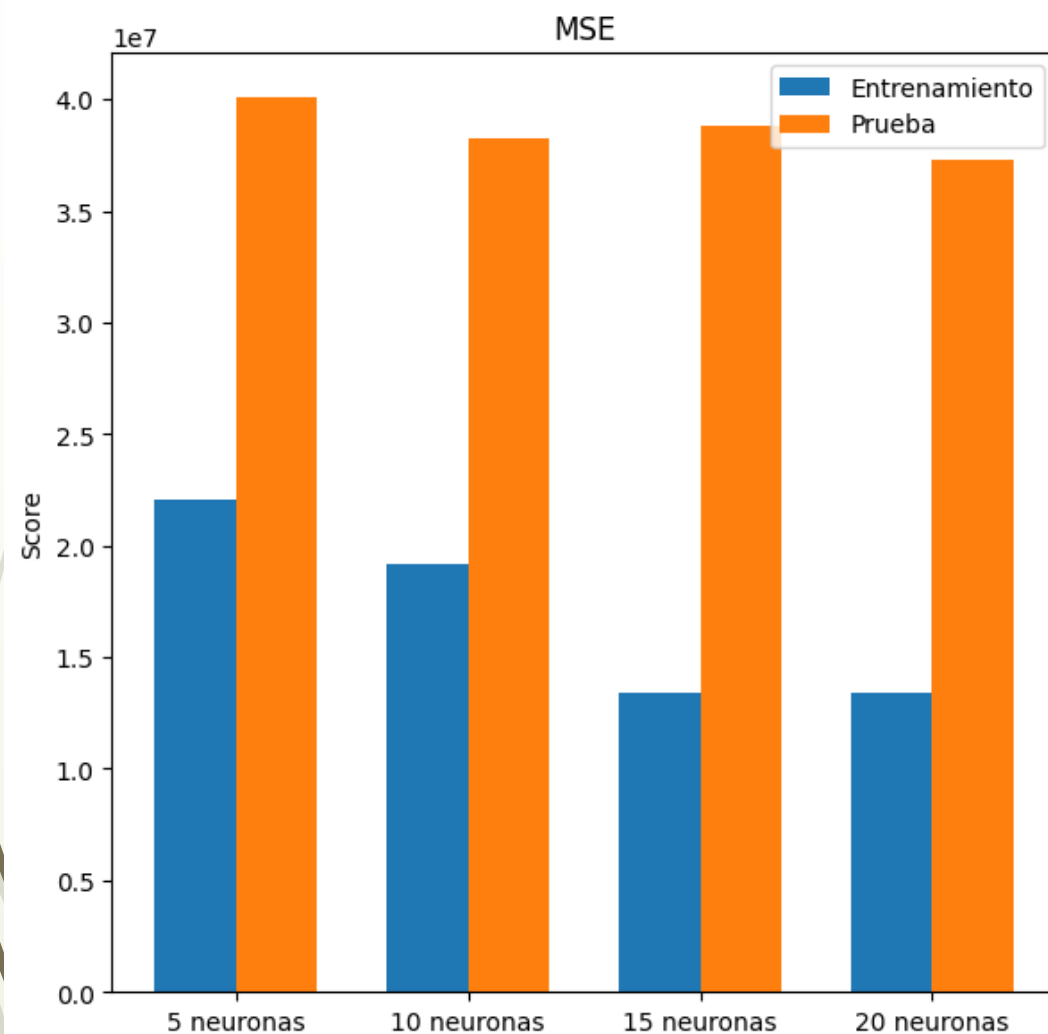
2. Red alimentada: 5,10,15 y **20 neuronas**



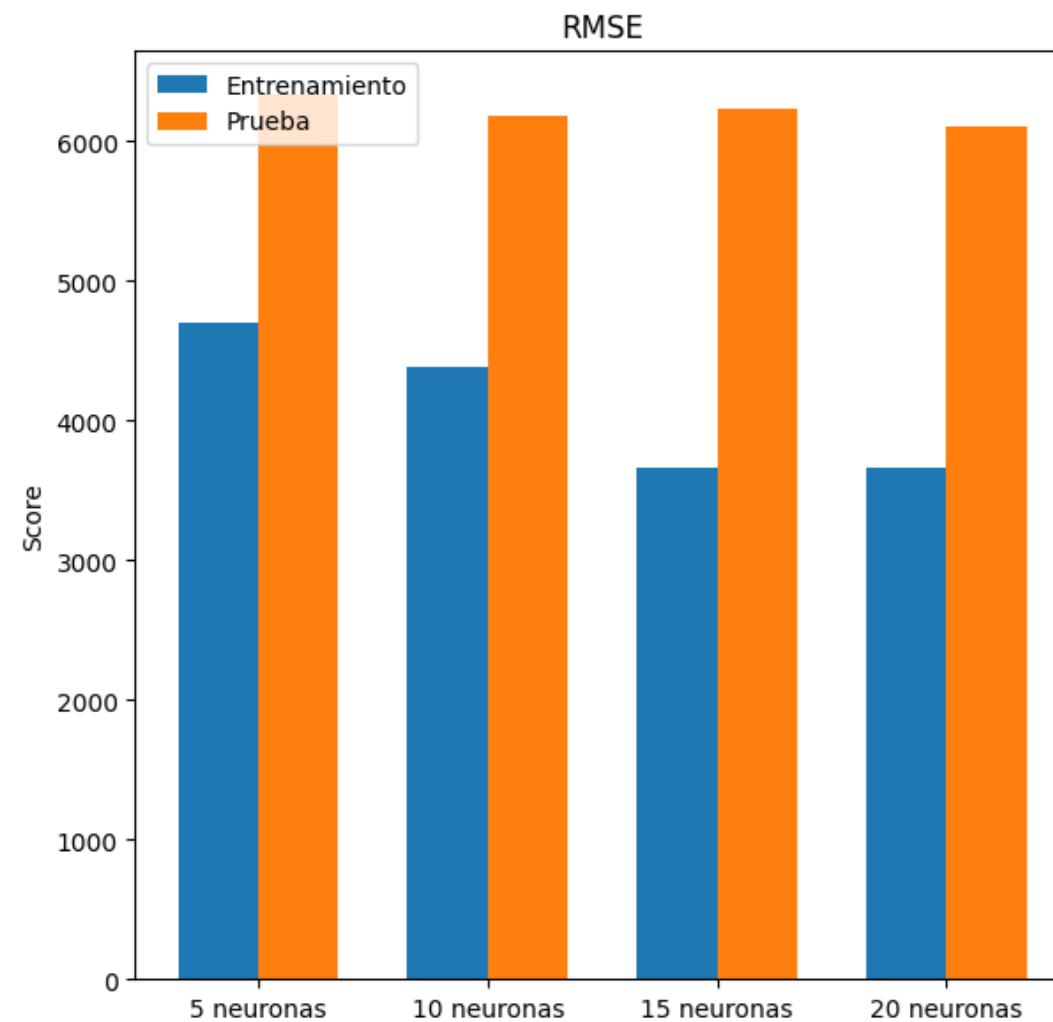
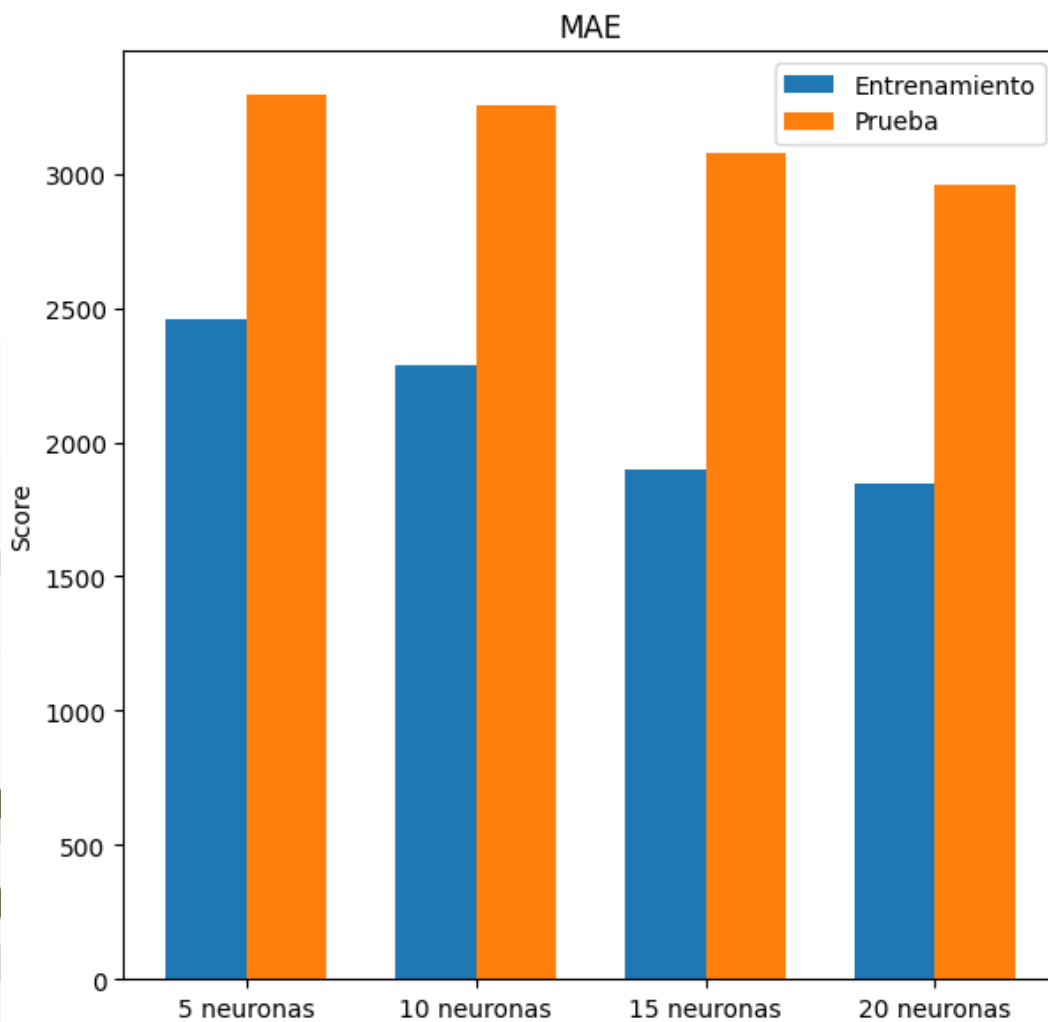
(Grecia) Comparación de los datos reales y los datos predichos (20 neuronas)



2. Red alimentada: 5,10,15 y 20 neuronas



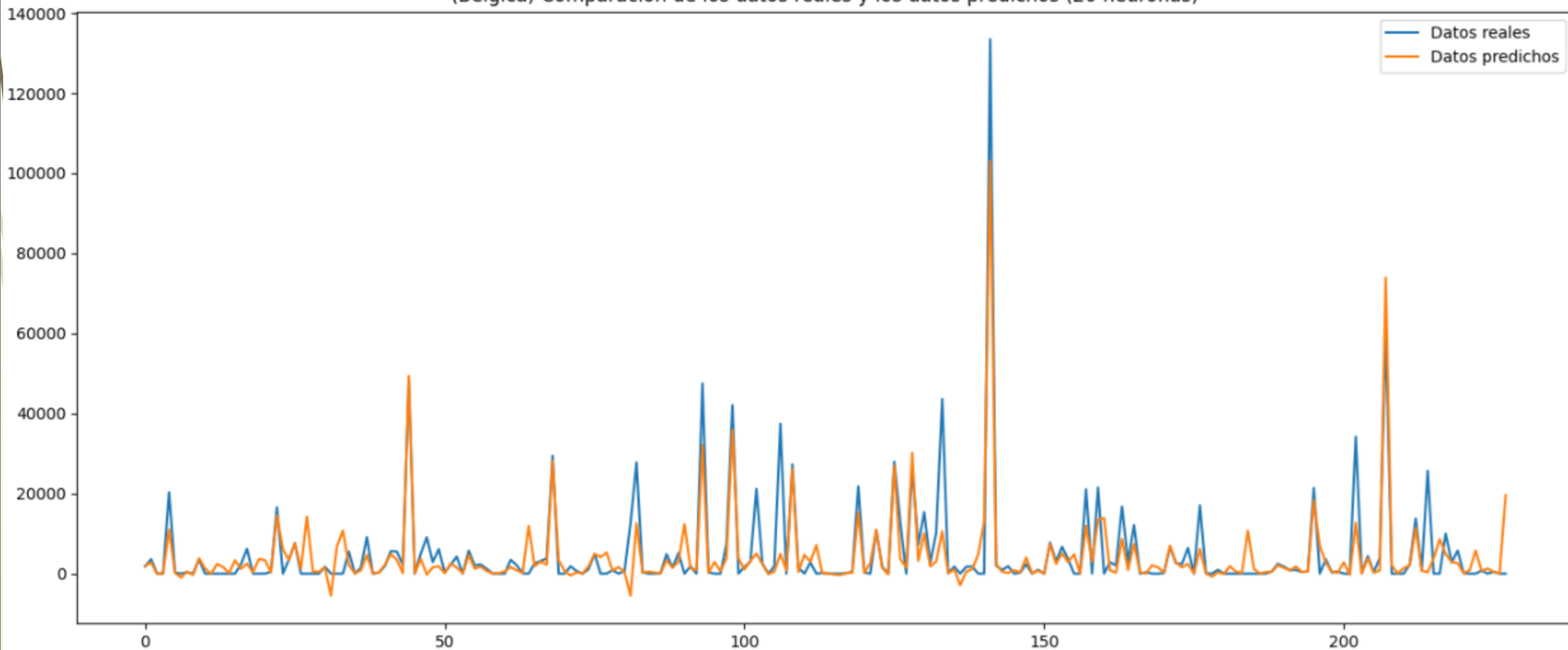
2. Red alimentada: 5,10,15 y 20 neuronas



2. Red alimentada: 5,10,15 y **20 neuronas**



(Bélgica) Comparación de los datos reales y los datos predichos (20 neuronas)



3. Mejor resultado: 1 o 2 capas de RNN



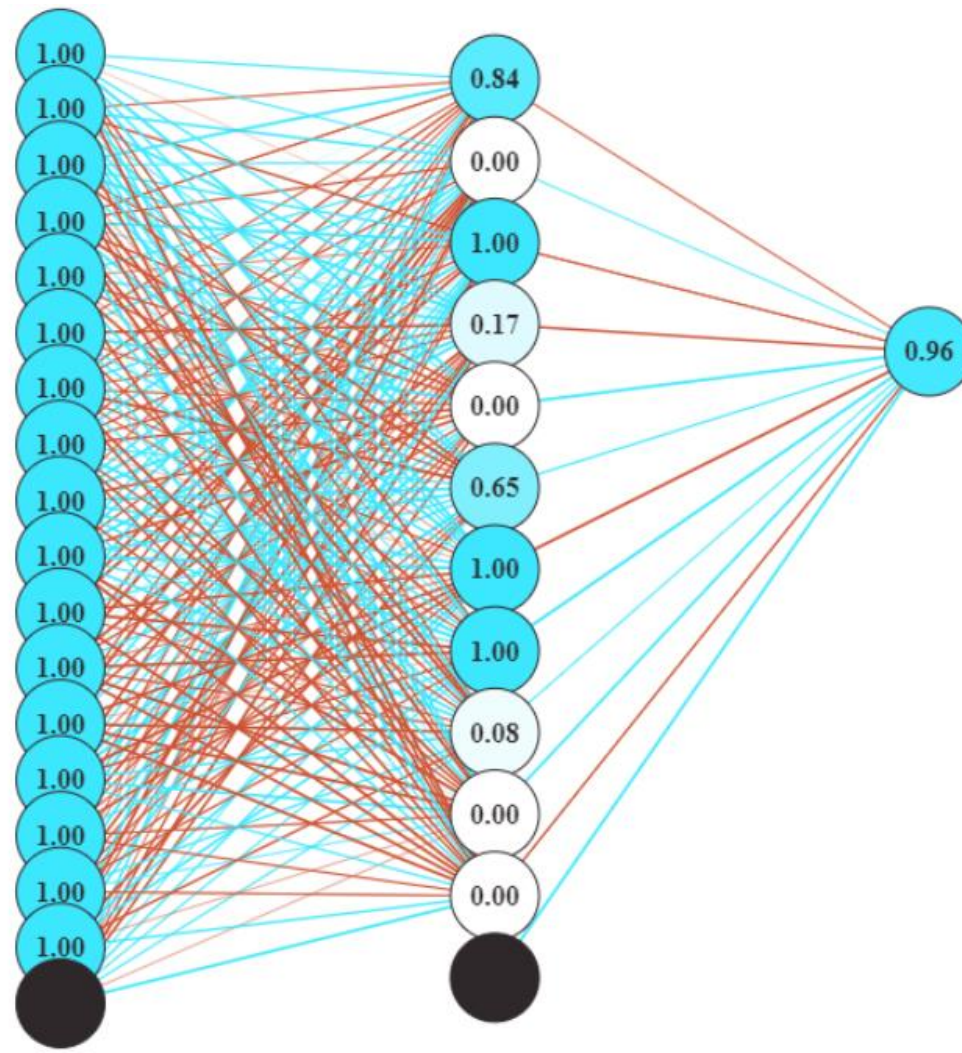
Mejores hiperparámetros encontrados: {'solver': 'lbfgs', 'momentum': 0.5, 'learning_rate': 'invscaling', 'hidden_layer_sizes': (17, 11)}

Mejor MSE de entrenamiento: 34333400.87595849

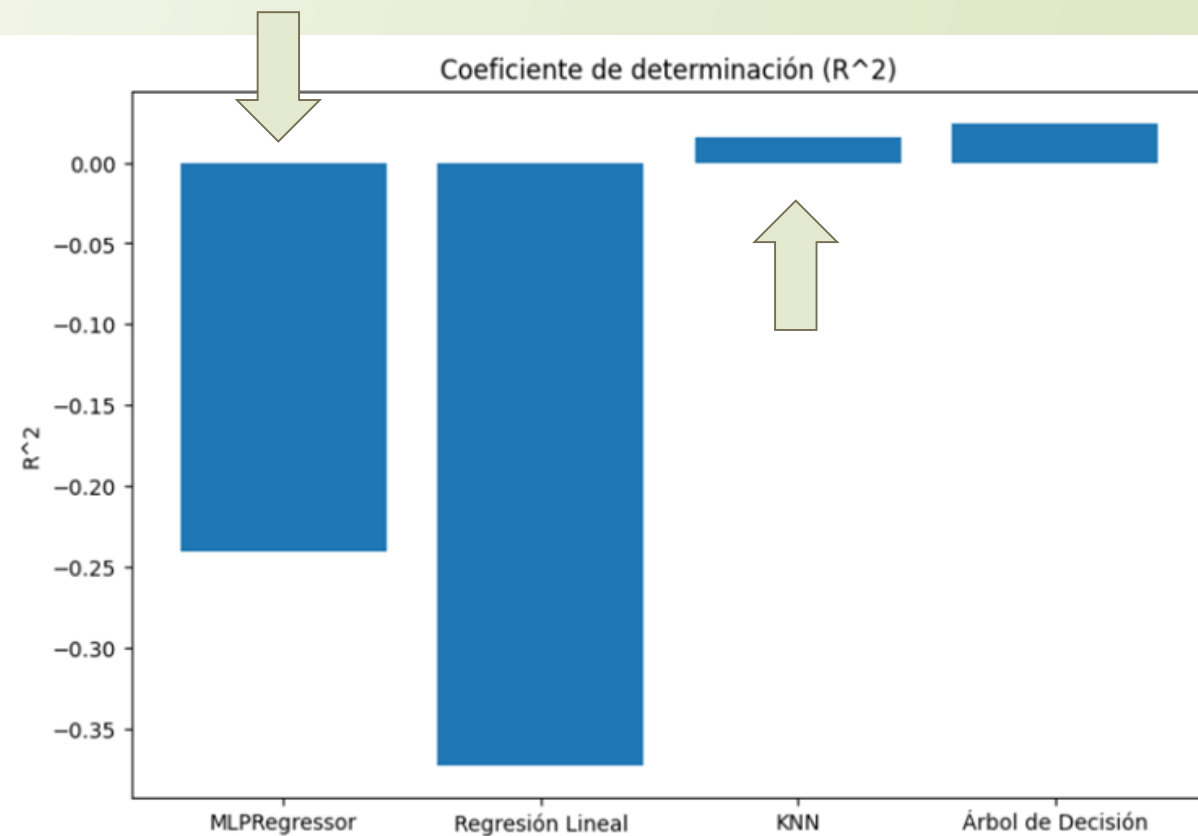
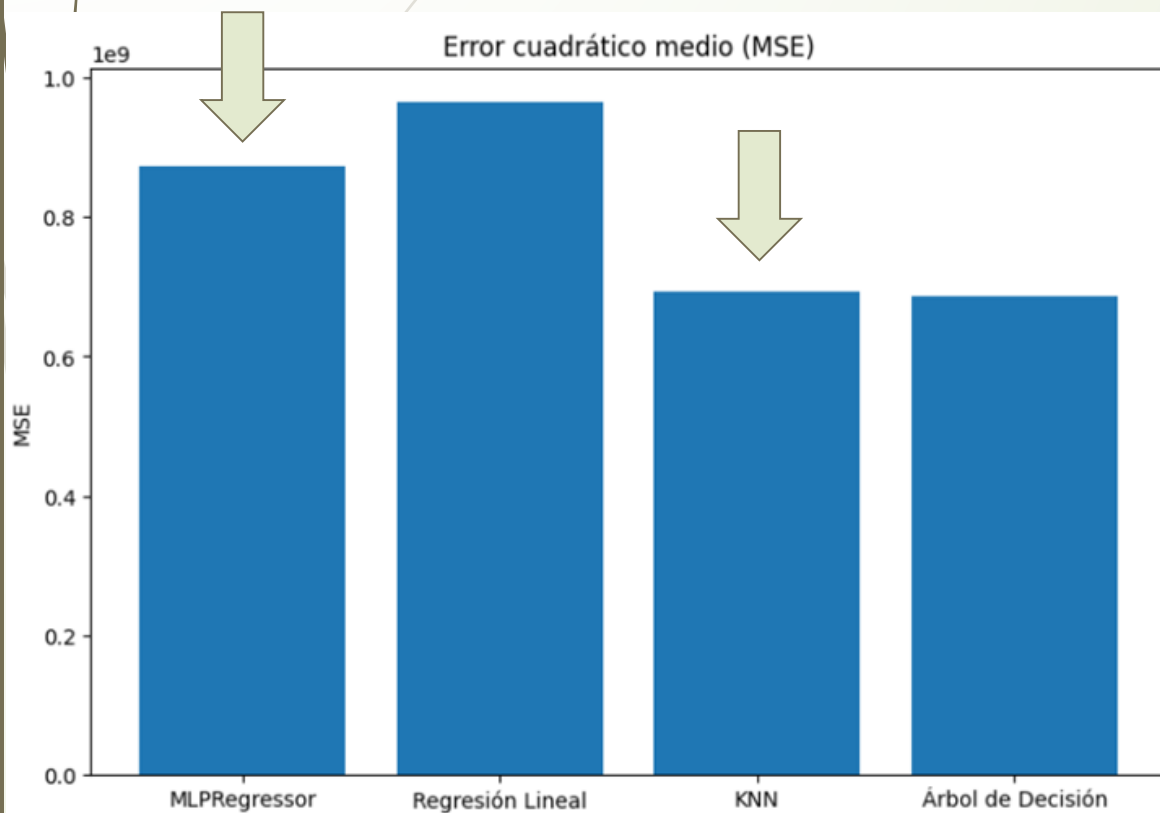
Mejor modelo:

MSE de prueba: 872900580.03448

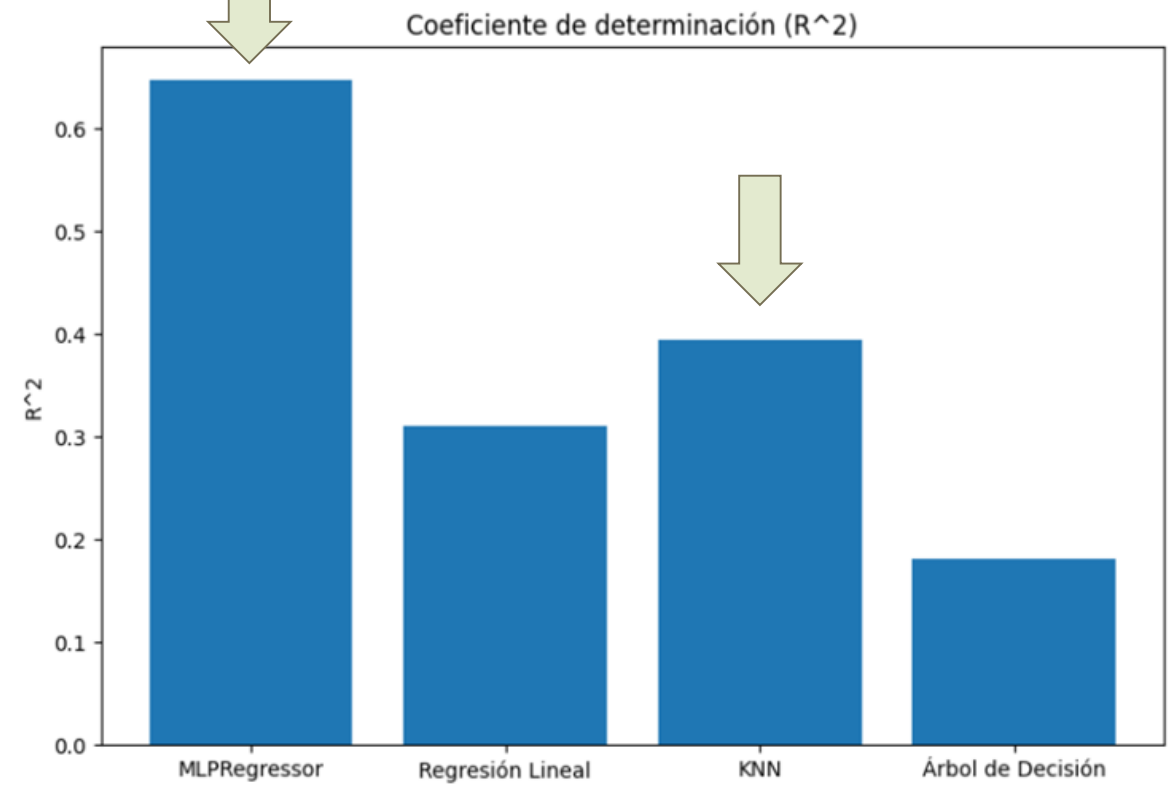
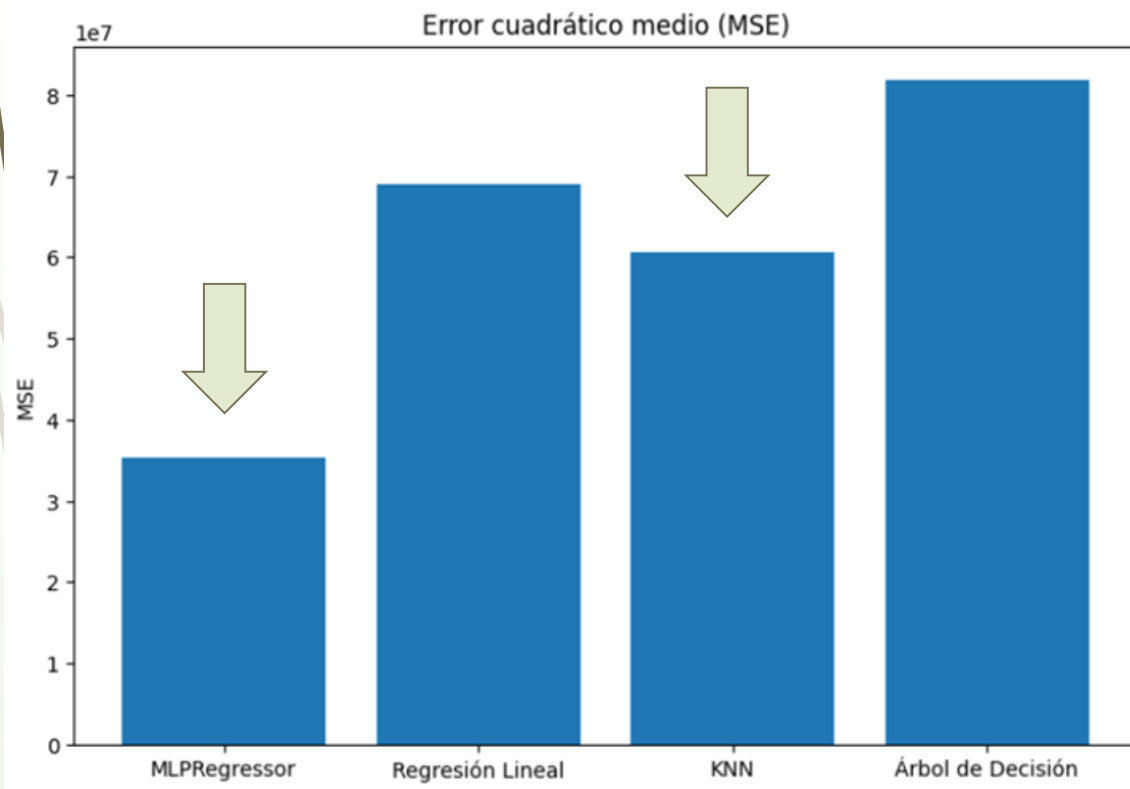
R2 de prueba: -0.24009



4. Comparación de modelos: lineal, knn, árbol y red neuronal



4. Comparación de modelos: lineal, knn, árbol y RNN



5. implementación de mejoras

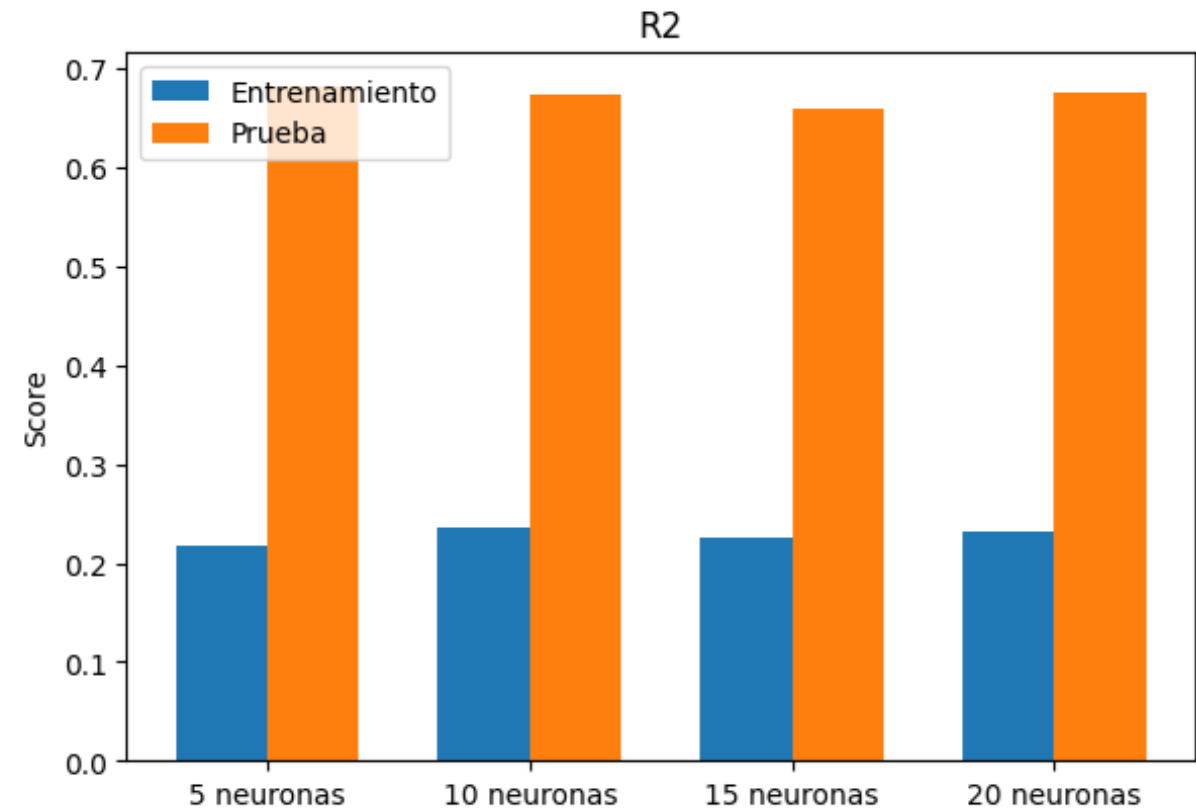
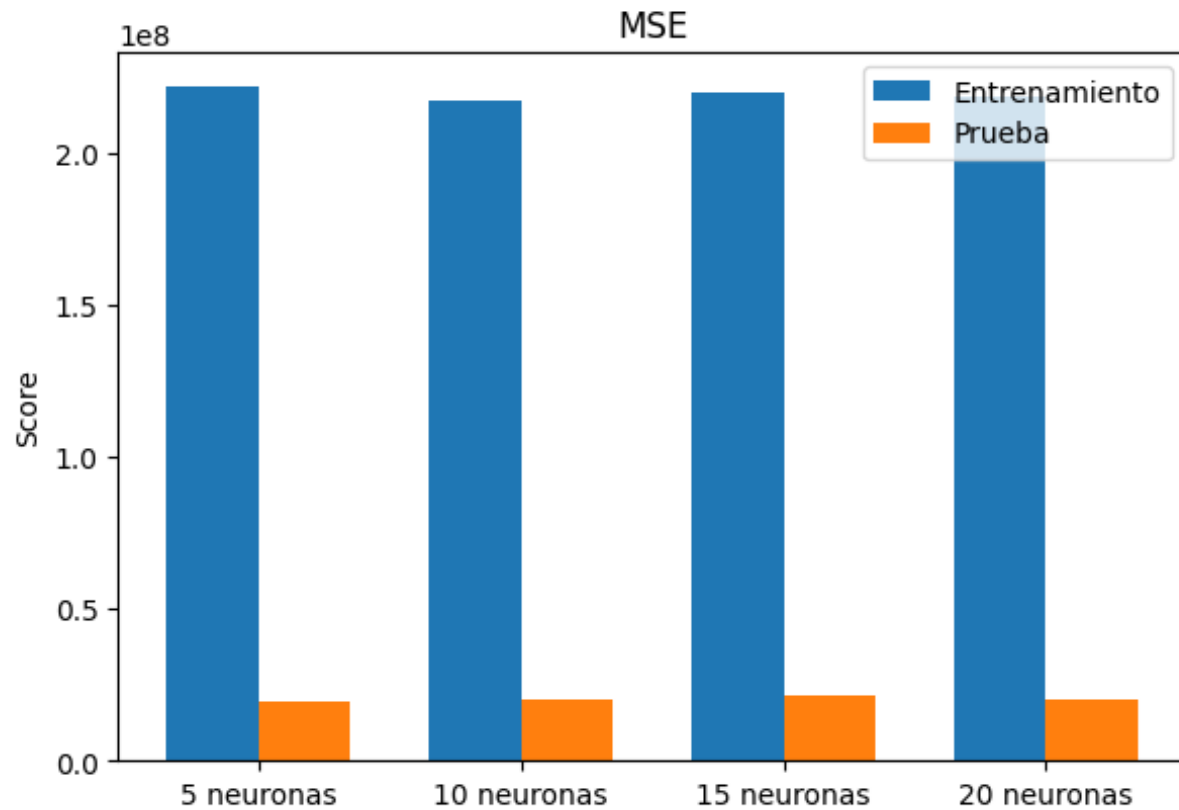
Nuestras mejoras son las siguientes principalmente:

Hemos decidido quitar los datos del principio ya que había cantidad de problemas a la hora de imputar los datos y realmente no nos los creemos mucho. empezamos a partir del **1 de marzo de 2020 en vez de 22 de enero de 2020**.

Por otro lado, como comentó el profesor en la clase del laboratorio, hemos querido implementar la matriz de días a nuestra matriz de datos. Básicamente tendríamos nuestra matriz de datos unida **con otra matriz de nº de la matrix x 7**. Sabiendo qué día fue datado entonces ponemos un 1 y en las demás columnas un 0, véase la tabla ejemplo:

Lunes	1	0	0	0	0	0	0
viernes	0	0	0	0	1	0	0
martes	0	1	0	0	0	0	0
jueves	0	0	0	1	0	0	0
lunes	1	0	0	0	0	0	0
sábado	0	0	0	0	0	1	0

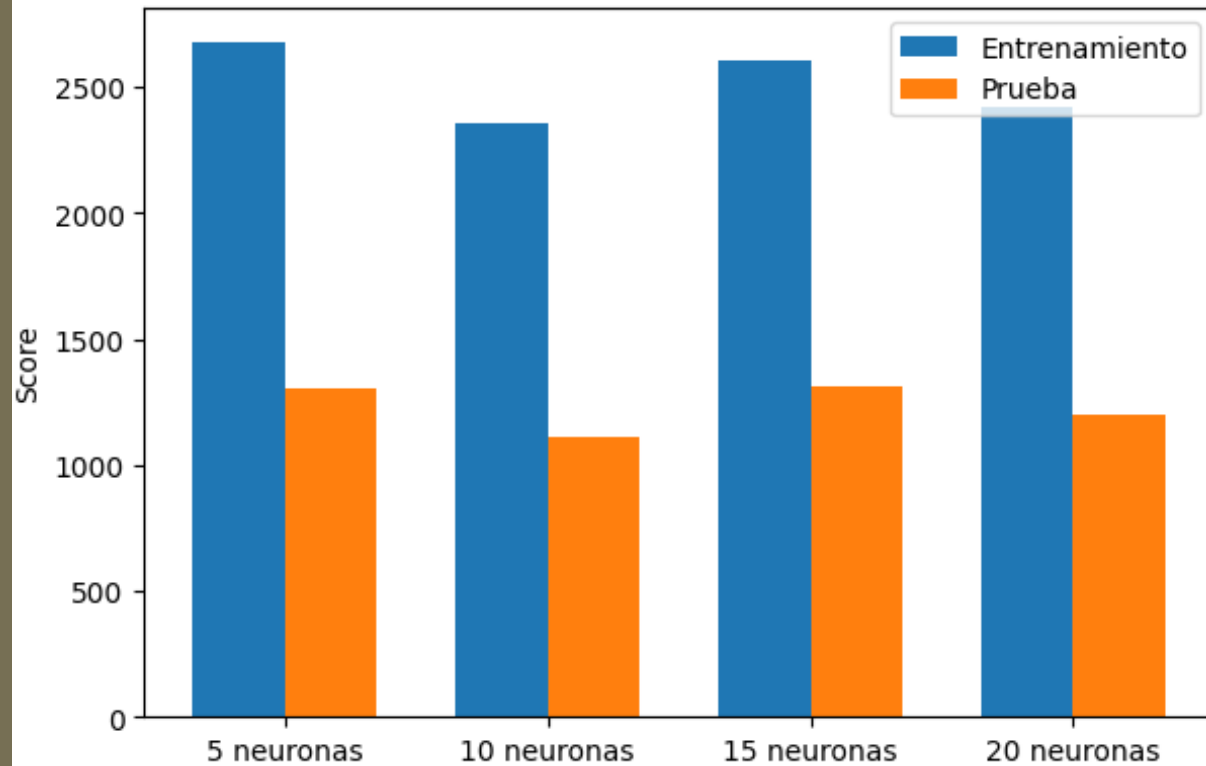
5. implementación de mejoras: 5,10,15 y 20 neuronas



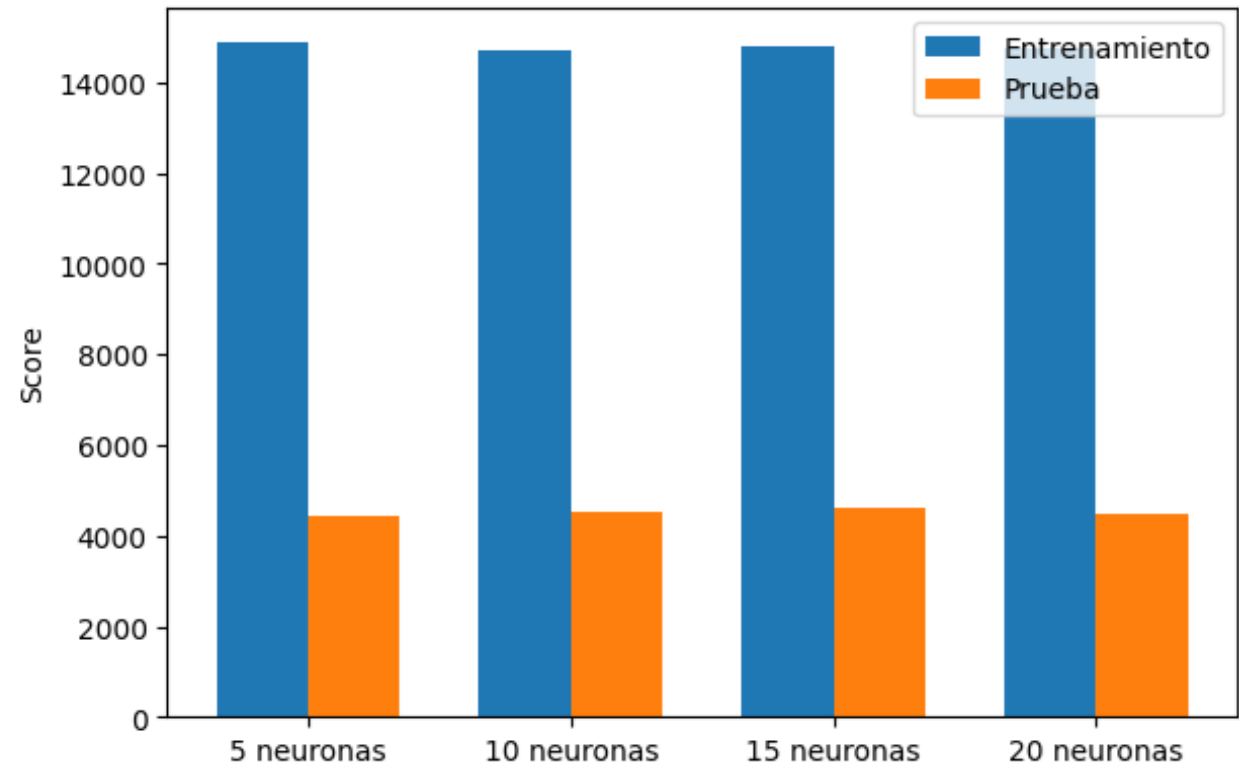
5. implementación de mejoras: 5,10,15 y 20 neuronas



MAE



RMSE

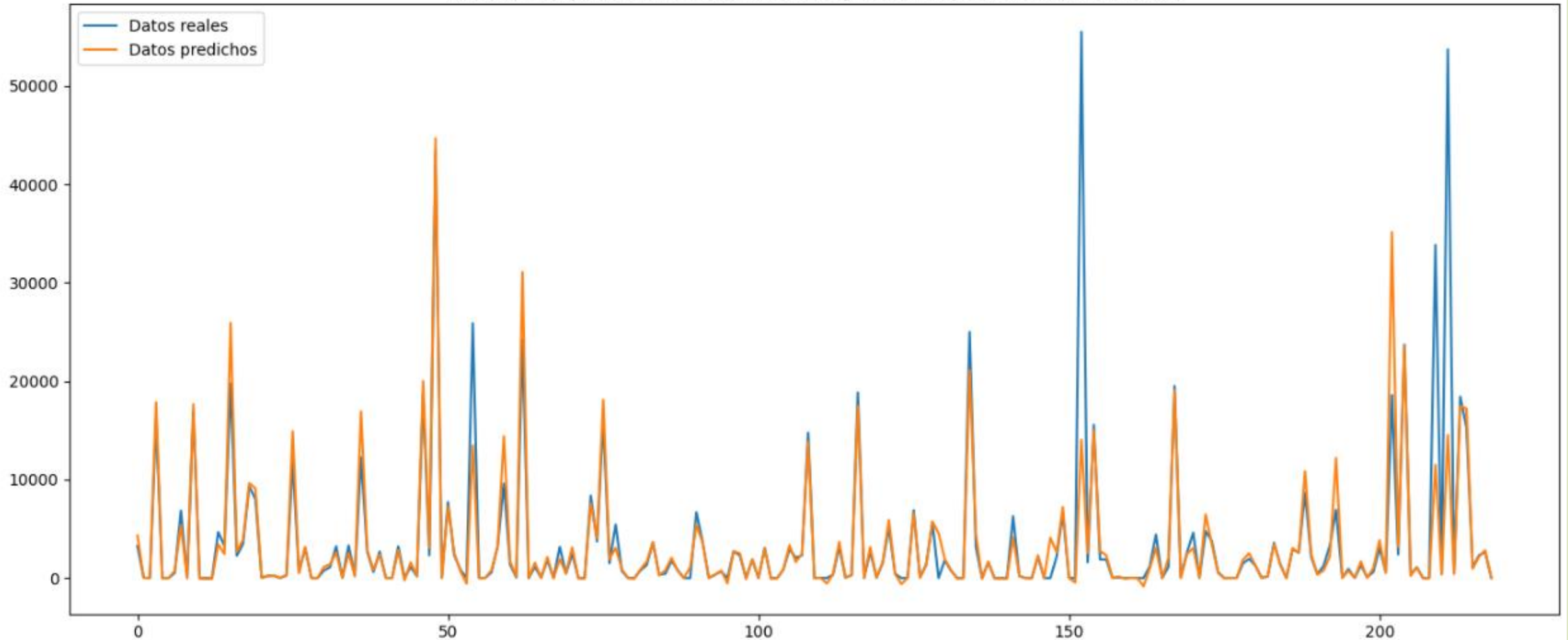


5. implementación de mejoras

10 neuronas



(Grecia) Comparación de los datos reales y los datos predichos (10 neuronas)

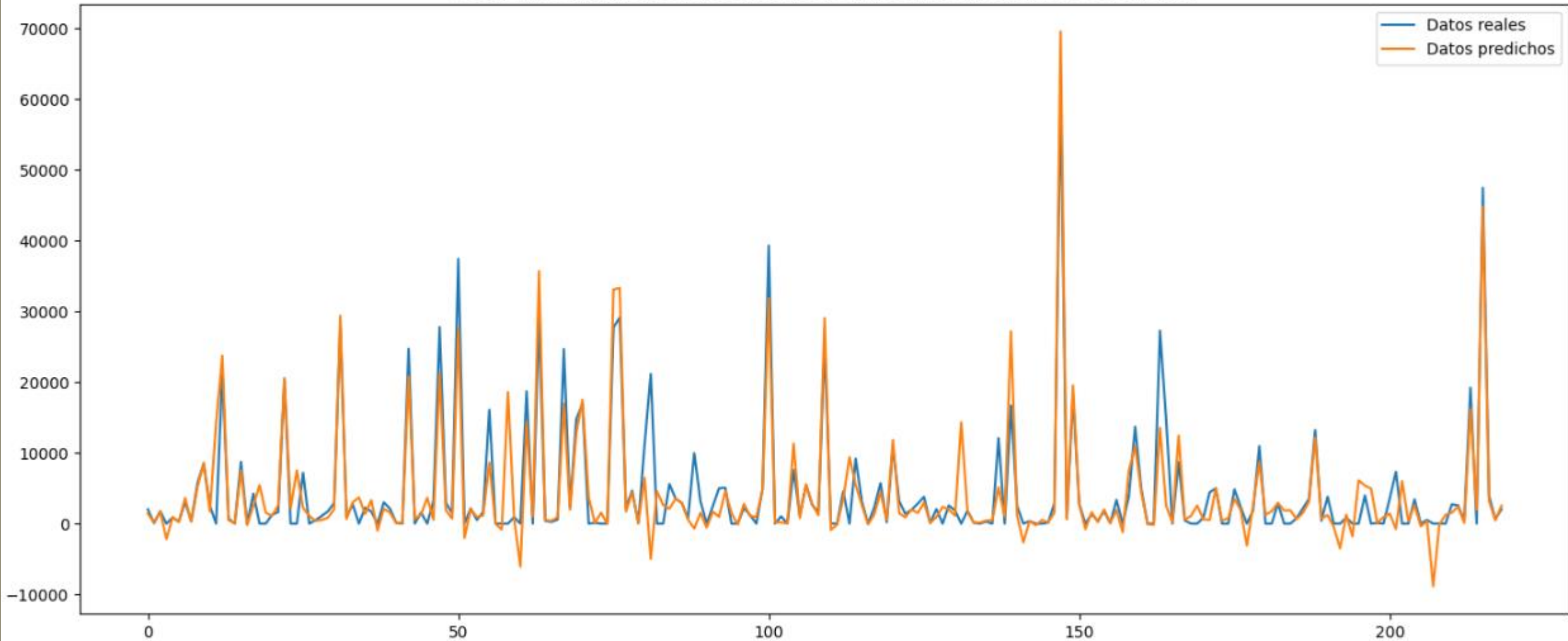


5. implementación de mejoras

20 neuronas



(Belgica) Comparación de los datos reales y los datos predichos (20 neuronas)





Conclusiones



Para el entrenamiento de modelos con RNN, **el preprocesado de los datos y la limpieza** de los mismos es importantísimo ya que si le consigues dar una matriz con mejores datos como en las mejoras implementadas será **capaz de trabajar mejor**.

Estos datos tienen que **ser relevantes con el significado del problema** que se está estudiando, **sino podrían tener un impacto negativo** y obtener peores resultados, al unir la matriz de los días a nuestra matriz de independientes se puede apreciar como en las métricas es mejor y por tanto predice mejor.