

Universidad de Costa Rica
Facultad de Ingeniería
Escuela de Ingeniería Eléctrica
Prof. Ing. Marvin Coto Jiménez

Proyecto 3

IE0435 - Inteligencia Artificial Aplicada a la Ingeniería Eléctrica

Red Neural MLP para predecir la Temperatura en la Estación Experimental Agrícola Fabio Baudrit Moreno de la Universidad de Costa Rica

Hecho por:

Fecha de entrega: 13 de Diciembre 2023

Kevin Delgado Rojas

E-Mail: kevin.delgadorojas@ucr.ac.cr

Carné: B82566

Resumen

Este proyecto consiste en la implementación de una red neuronal del tipo Perceptrón Multicapa (MLP, por sus siglas en inglés) para realizar predicciones de valores de temperatura en la Estación Experimental Agrícola Fabio Baudrit Moreno. La red neuronal utilizada posee una arquitectura compuesta por 50 capas y se entrenó durante 50 épocas.

Los resultados obtenidos del modelo de predicción fueron los siguientes:

- MSE (Error Cuadrático Medio): 0.1029
- RMSE (Raíz del Error Cuadrático Medio): 0.3208
- MAE (Error Absoluto Medio): 0.2276

Estos resultados se utilizan para evaluar el rendimiento del modelo de predicción. El MSE, RMSE y MAE son métricas que indican cuánto se desvían las predicciones del modelo de los valores reales. En este caso, los valores obtenidos indican que las predicciones del modelo tienen errores promedio en torno a 0.1 a 0.32 grados Celsius en relación con los valores reales.

Índice

Índice de figuras	III
Índice de cuadros	IV
1 Introducción	1
2 Procedimiento	2
2.1 Solicitud de datos climáticos	2
2.2 Visualización de datos	2
2.3 Relleno de datos de tiempo	3
2.4 Relleno de datos de temperatura	4
2.5 Implementación de Red Neuronal MLP	5
3 Resultados	6
4 Conclusiones	9
5 Recomendaciones	9
6 Referencias	10

Índice de figuras

Figura 1:	Gráfico de diagnóstico de datos	2
Figura 2:	Cantidad de filas original	3
Figura 3:	Cantidad de filas nuevas	3
Figura 4:	Promedios de temperatura	4
Figura 5:	Gráfica del conjunto de datos sin sesgos	5
Figura 6:	Comparación de datos reales vs predicciones	6
Figura 7:	Caption	7
Figura 8:	Comparación de gráficas	8

Índice de cuadros

Tabla 1:	Métricas de evaluación del modelo	6
----------	---	---

1. Introducción

La predicción de series temporales se ha convertido en un pilar fundamental para anticipar eventos futuros basados en datos históricos. En este contexto, el uso de redes neuronales, como las Multilayer Perceptrons (MLP) y las Recurrent Neural Networks (RNN), ofrece un enfoque robusto y dinámico para predecir valores climáticos como la temperatura. Este proyecto se enfoca en aplicar estas redes neurales para prever la temperatura en la Estación Experimental Agrícola Fabio Baudrit Moreno de la Universidad de Costa Rica.

La predicción de series temporales implica la extensión de datos históricos hacia el futuro, donde aún no se han registrado mediciones[2]. Para lograr un pronóstico preciso, se deben considerar variables cruciales como la cantidad de períodos históricos y el horizonte de predicción. Este enfoque no solo es relevante para la meteorología, sino que también encuentra aplicaciones fundamentales en diversas industrias. Entre estas aplicaciones, una de las más destacadas es la capacidad de anticiparse a las necesidades del mercado mediante la predicción de la demanda[1].

En este contexto, el estudio de predicción de la temperatura en la Estación Experimental Agrícola no solo representa un desafío técnico, sino que también tiene implicaciones prácticas significativas. El empleo de redes neuronales MLP permitiría predecir temperaturas, lo cual es importante para la toma de decisiones y la planificación de las actividades agrícolas del lugar.

Así, este proyecto busca no solo explorar el rendimiento de las redes neuronales en la predicción de temperaturas, sino también demostrar su viabilidad en la generación de pronósticos confiables, con la intención de ofrecer herramientas predictivas útiles para la gestión de recursos y la toma de decisiones en el ámbito agrícola.

2. Procedimiento

2.1. Solicitud de datos climáticos

La Estación Experimental Agrícola Fabio Baudrit Moreno de la Universidad de Costa Rica cuenta con una plataforma llamada Netbeat, donde se registran diariamente las temperaturas del sitio que son medidos cada 15 minutos. Esta plataforma no es muy antigua, ya que apenas se instaló en Enero del año pasado, sin embargo cuenta con una buena cantidad de datos para el desarrollo del proyecto.

2.2. Visualización de datos

Con el fin de tener un diagnóstico y conocer el conjunto de datos, se gráfico el historial de datos, como se muestra en la siguiente figura 1.

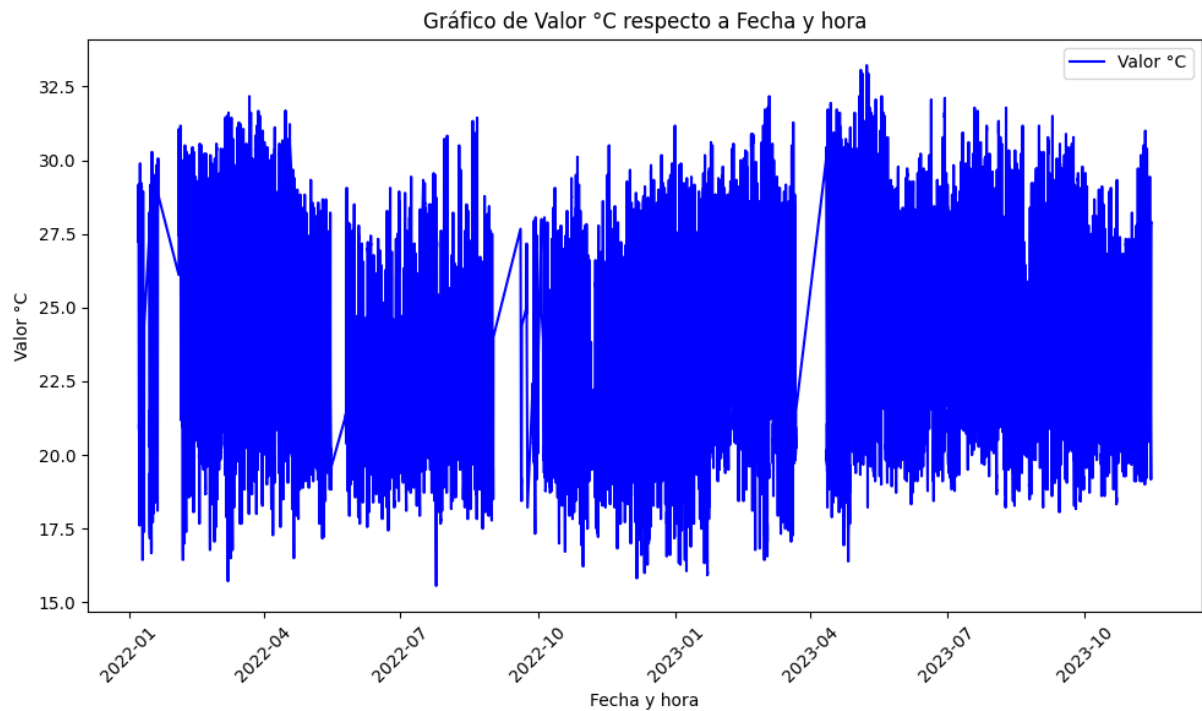


Figura 1: Gráfico de diagnóstico de datos

Como se aprecia en la figura anterior, existen sesgos significativos en el conjunto de datos, lo cual implica un problema. Al investigar un poco mas a fondo el conjunto de datos, se observó que habían faltantes de datos horas específicas y varios días. Pudo ser que el sistema de sensores fallara o desconectarán los dispositivos.

2.3. Relleno de datos de tiempo

Es necesario aplicar técnicas para tener esos datos faltantes y poder tener el conjunto de datos completo. En un inicio el conjunto de datos posee la siguiente cantidad de filas: 54873 filas.

	Fecha y hora	Valor °C	Nombre del grafico
0	2023-11-14 11:11:27	27.89	UCR - Out emp
1	2023-11-14 10:56:18	27.89	UCR - Out emp
2	2023-11-14 10:41:09	27.83	UCR - Out emp
3	2023-11-14 10:26:00	27.56	UCR - Out emp
4	2023-11-14 10:10:50	27.39	UCR - Out emp
...
54868	2022-01-07 11:36:21	27.22	UCR - Out emp
54869	2022-01-07 11:21:08	28.17	UCR - Out emp
54870	2022-01-07 11:05:57	28.56	UCR - Out emp
54871	2022-01-07 10:50:49	28.11	UCR - Out emp
54872	2022-01-07 10:35:39	27.28	UCR - Out emp
[54873 rows x 3 columns]			

Figura 2: Cantidad de filas original

Primeramente, se comenzó con el relleno de datos de tiempo, por lo cual se aplicó la función llamada:

```
def encontrar_lapsos_mayores1(df, inicio, fin, tiempo_minutos)
```

Para utilizar esta función primero se buscó las fechas donde estaban los faltantes, por eso recibe como parámetros una fecha de inicio, final y la cantidad de minutos que queramos agregar. Aplicando la función se logró rellenar el conjunto de datos, llenado a la siguiente cantidad de filas: 62369 filas.

	Fecha y hora	Valor °C	Nombre del grafico
0	2022-01-07 10:35:39	27.28	UCR - Out emp
1	2022-01-07 10:50:49	28.11	UCR - Out emp
2	2022-01-07 11:05:57	28.56	UCR - Out emp
3	2022-01-07 11:21:08	28.17	UCR - Out emp
4	2022-01-07 11:36:21	27.22	UCR - Out emp
...
62327	2023-11-14 10:10:50	27.39	UCR - Out emp
62328	2023-11-14 10:26:00	27.56	UCR - Out emp
62329	2023-11-14 10:41:09	27.83	UCR - Out emp
62330	2023-11-14 10:56:18	27.89	UCR - Out emp
62331	2023-11-14 11:11:27	27.89	UCR - Out emp
[62369 rows x 3 columns]			

Figura 3: Cantidad de filas nuevas

2.4. Relleno de datos de temperatura

Esta fue la parte más difícil del problema y la que conllevó más tiempo. Para añadir datos de temperatura existen varias técnicas de interpolación que consisten en calcular un valor según los puntos que se tengan. Sin embargo, al existir sesgos tan grandes, las técnicas no funcionaban porque agregaban valores muy grandes, ya que seguían de manera lineal.

La mejor solución al problema, que conllevó un costo significativo de tiempo fue encontrar el promedio de la temperatura al inicio, medio y fin de un día. Por lo cual, se calculó, y se obtuvo lo siguiente:

```
Promedio de temperatura al inicio del día: 20.711305057096247
Promedio de temperatura al final del día: 20.77194127243067
Promedio de temperatura al mediodía: 28.07015517241379
```

Figura 4: Promedios de temperatura

Se tomaron esos datos y a mano se fueron agregando en el conjunto de datos, ya que era difícil añadirlos de manera automática, porque los datos de tiempo no eran exactos, lo cual lo dificultaba.

Posteriormente, gracias a la ayuda del estudiante Juan Ángel, me recomendó utilizó la técnica de rellenar con la recta, por lo cual se hizo la siguiente función:

```
def rellenar_con_recta(dataframe, columnas_a_rellenar)
```

La cual recibe un dataframe, y le decimos que columna queremos rellenar, en este caso sería la columna de Valor. Una vez hecho esto se obtuvo el nuevo conjunto de datos con las temperaturas, como se muestra en la siguiente figura 5.

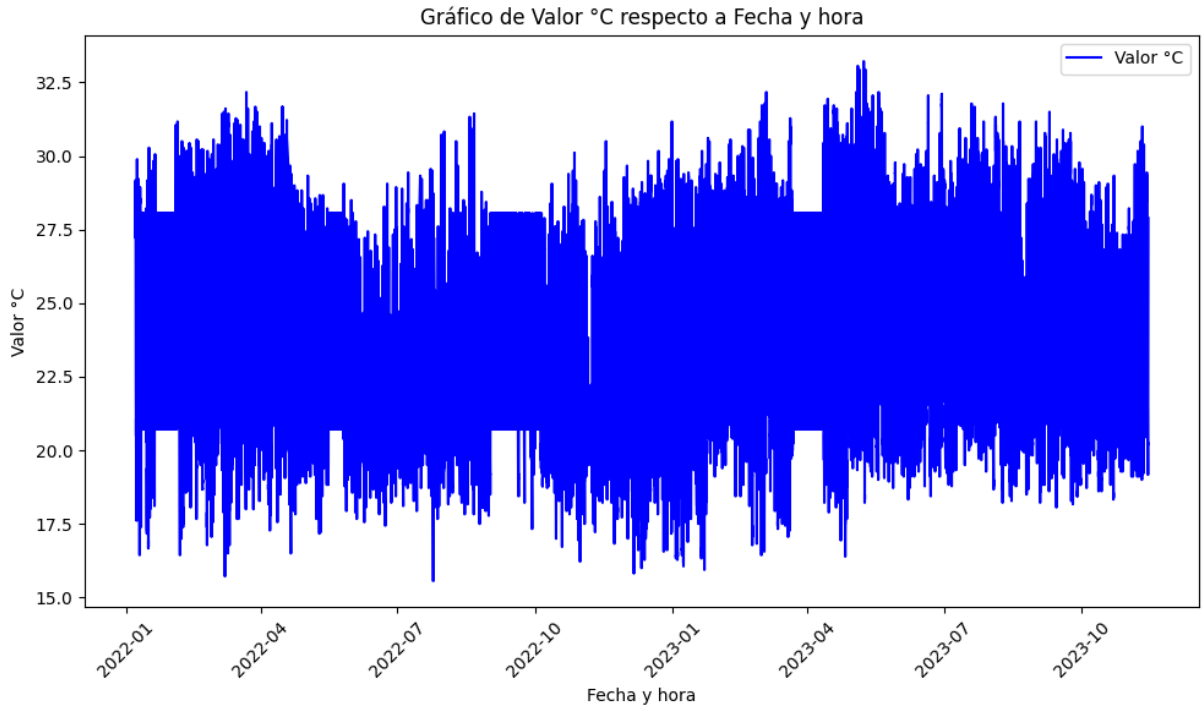


Figura 5: Gráfica del conjunto de datos sin sesgos

Como se observa en la figura anterior, ya no existen los sesgos, sin embargo, al haber añadido los promedios de la temperatura, hace que no se vea el comportamiento aleatorio que nos gustaría. Pero para este problema se va a proceder a trabajar con el nuevo conjunto de datos.

2.5. Implementación de Red Neuronal MLP

Para este proyecto se van implementar el tipo de red neuronal Multilayer Perceptrons (MLP), de 50 capas y 50 épocas para la actualización de los pesos. Se tomó esta decisión debido a que el conjunto de datos no es muy grande. Con respecto al conjunto de entrenamiento se eligió un 90 % de entrenamiento y 10 % de prueba, para tener resultados mas precisos, ya que se trata de temperaturas. La implementación de las mismas se hará en un google collab, el cual estará disponible en la entrega de este informe.

3. Resultados

Con respecto a los resultados alcanzados, la red MLP se comportó de buena manera, consiguiendo los siguientes resultados.

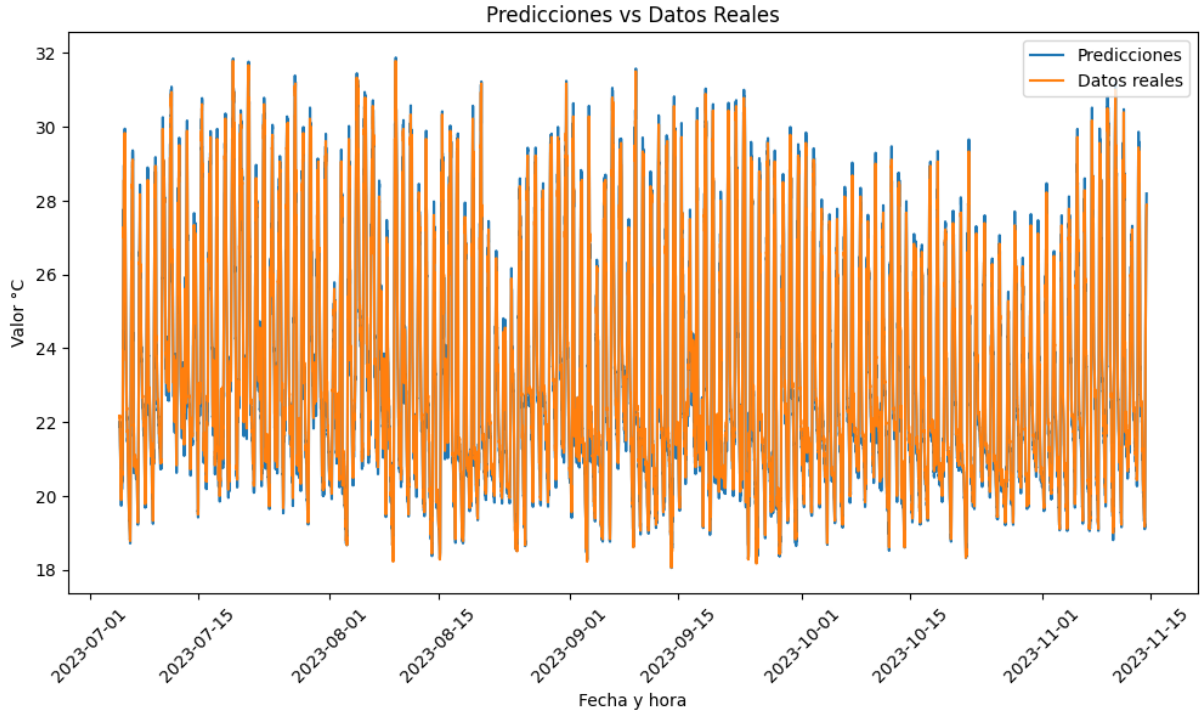


Figura 6: Comparación de datos reales vs predicciones

Como se muestra en la figura 6, al comparar se puede visualizar que el modelo, logró predecir los datos de temperatura para el conjunto de prueba. Se puede notar que los datos de temperatura predichos por el modelo, se acercan en gran parte a los datos reales.

Con respecto a las métricas de evaluación, se obtuvieron los siguientes resultados:

Métrica	Valor
MSE	0.10291
RMSE	0.3208
MAE	0.2276

Cuadro 1: Métricas de evaluación del modelo

1. MSE (Error Cuadrático Medio): Es la media de los errores al cuadrado entre las predicciones del modelo y los valores reales. Un MSE más bajo indica que el modelo tiene menores errores en sus predicciones. En este caso, el MSE de aproximadamente 0.1029 indica que, en promedio, las predicciones del modelo están a una distancia considerablemente baja (en términos cuadráticos) de los valores reales.

2. RMSE (Raíz del Error Cuadrático Medio): Es simplemente la raíz cuadrada del MSE y proporciona una métrica del error similar al MSE pero en la misma escala que los datos originales. En este caso, el RMSE de aproximadamente 0.3208 sugiere que las predicciones del modelo tienen un error promedio (en la misma escala que los datos originales) de alrededor de 0.3208 grados Celsius respecto a los valores reales.
3. MAE (Error Absoluto Medio): Es la media de los errores absolutos entre las predicciones del modelo y los valores reales. Es menos sensible a valores atípicos en comparación con el MSE, ya que no considera los errores al cuadrado. En este caso, un MAE de aproximadamente 0.2276 indica que, en promedio, las predicciones del modelo difieren en aproximadamente 0.2276 grados Celsius de los valores reales.

Así mismo, se graficó la validación del modelo, obteniendo los siguientes resultados. Se puede mencionar que los picos abruptos cerca de la época 20 pueden ser indicios de inestabilidad del modelo, sin embargo, estos se mejoran conforme pasan las épocas de entrenamiento. Como se muestra en la figura 7.

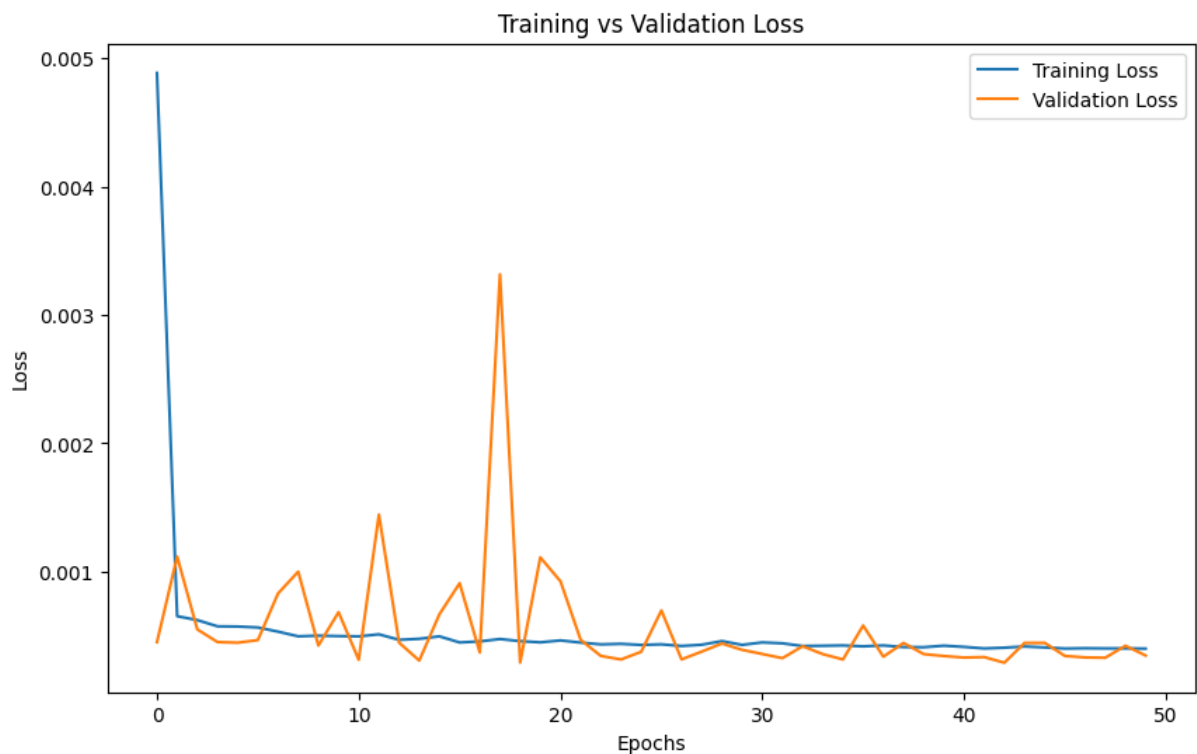


Figura 7: Caption

Por último, se graficó de manera independiente las predicciones hechas por el modelo, donde se aprecia de mejor manera el comportamiento.

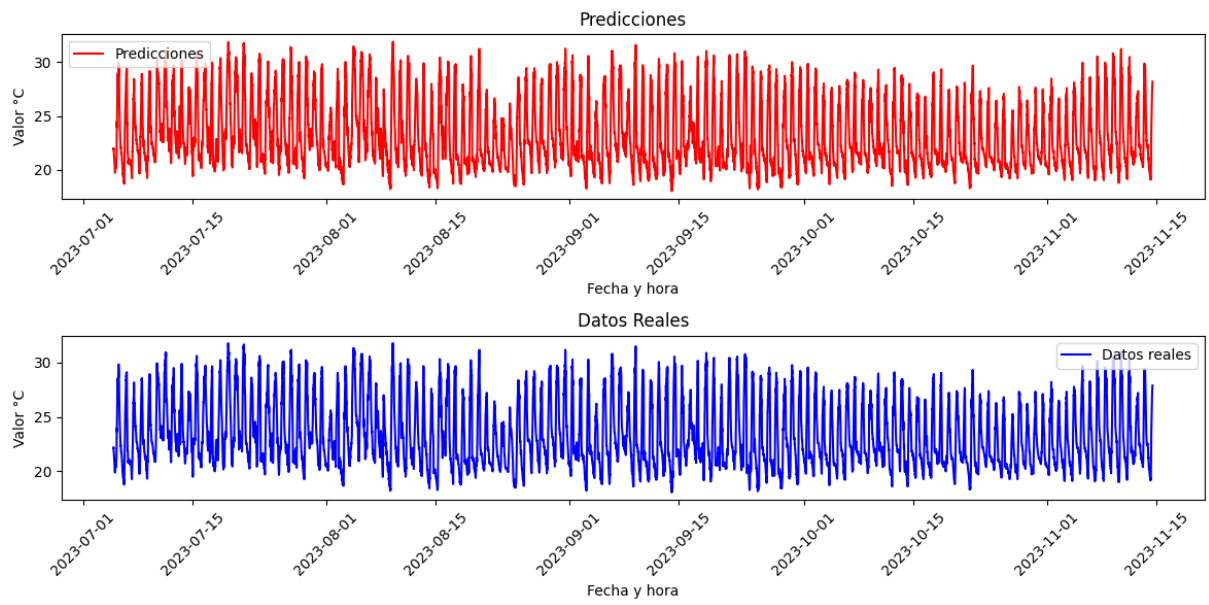


Figura 8: Comparación de gráficas

4. Conclusiones

El empleo de redes neuronales, particularmente Multilayer Perceptrons (MLP) para predecir valores de temperatura en la Estación Experimental Agrícola Fabio Baudrit Moreno de la Universidad de Costa Rica ha demostrado ser una estrategia eficaz y prometedora, y con mas tiempo puede mejorarse en gran medida.

Precisión en las predicciones: Se observó que las red neuronal MLP logró unas métricas de evaluación bastante acertadas. Esto permitió generar pronósticos con una precisión notable.

Las predicciones precisas de temperatura tienen un impacto significativo en la toma de decisiones agrícolas para los estudiantes de agronomía. Este modelo podría ser empleado para planificar estrategias de cultivo, gestionar recursos hídricos, entre otros.

A pesar de los resultados positivos, es esencial realizar una evaluación continua y ajustar mejor el modelo para tener una mejor precisión y adaptabilidad a posibles cambios en los datos o en el entorno.

5. Recomendaciones

Los sesgos en los conjuntos de datos pueden ser un gran problema a la hora de trabajarlos, en este caso los de temperatura consumieron una cantidad de tiempo considerable, por lo cual es mejor conocer el conjunto de datos.

Ajustar de una mejor manera el modelo empleado y utilizar distintas capas, pueden ayudar a que las predicciones sean más precisas.

Se puede implementar otro tipo de Red Neural como RNN para predecir con el mismo conjunto de datos, para este proyecto no se hizo por cuestiones de tiempo, pero sería una alternativa para verificar cual predice mejor.

6. Referencias

Referencias

- [1] Cyberclick. «Data Science: predicciones de series temporales con machine learning.» (2023), dirección: <https://www.cyberclick.es/numerical-blog/data-science-predicciones-de-series-temporales-con-machine-learning>.
- [2] PRICING. «Series de Tiempo.» (2023), dirección: <https://www.pricing.cl/conocimiento/series-de-tiempo/#:~:text=La%20predicci%C3%B3n%20de%20series%20de,per%C3%ADodos%20y%20horizonte%20de%20predicci%C3%B3n..>