
# PROYECTO FINAL

# INTELIGENCIA ARTIFICIAL

**ALEXANDRA BARRIOS ROJAS**

**EMANUEL PALACIO URREGO**

**JUAN MANUEL MESA HENAO**

## 1. Introducción

La energía eléctrica se ha convertido en un bien fundamental para el desarrollo de la vida humana, esto lo podemos evidenciar con el crecimiento en la demanda de este bien que hemos visto a lo largo de este siglo, además la demanda de energía eléctrica de un país es considerada como un indicador de desarrollo por varias organizaciones. De ahí a que el sector eléctrico sea un gran negocio de interés, tanto para agentes privados como públicos, y es en ese punto donde se hace importante la regulación por parte de los países, pues ya que es importante brindar el precio adecuado a los usuarios, asegurando así la accesibilidad a este servicio, que actualmente es uno de los objetivos de desarrollo sostenible. Acá cobra importancia el hecho de vigilar los precios y saber cuál es la tendencia de estos, ya que por temas geopolíticos este precio está sometido a una gran incertidumbre y se hace importante para los países poder tomar decisiones adecuadas, en este caso con base en modelos. Por eso para este trabajo se ha elegido el precio de la electricidad como tema de interés.

Para este caso se decidió estar en dos escenarios diferentes, primero como un comercializador, al que le interesa comprar cuando sea barato, y vender cuando sea costoso, para este caso se realizó un modelo que predijera si el siguiente valor iba a bajar o a subir. Y para el otro escenario como la comisión de regulación de un país, a la que le interesa tener una predicción a más largo plazo, ósea tener un modelo que ele pueda ingresar datos de ciertos escenarios que puedan ocurrir, como un fenómeno del niño, por ejemplo, y tomar decisiones en base a estas predicciones, decisiones que pueden ir desde comprar más petróleo, embalsar más agua, o crear nuevas plantas, etc.

### Data Set utilizado

El data set utilizado, se encontró en la plataforma kaggle: https://www.kaggle.com/datasets/nicholasjhana/energy-consumption-generation-prices-and-weather

Contexto de los datos: el data set usado es del sistema eléctrico español, en este se encuentran diferentes tipos de generación (fossil oil, wind, hydro, etc) y su aporte en KW/h, se tiene la demanda prevista, y la demanda real y por último se tiene el precio en euros del KW/h, esto se encuentra por cada hora, durante 4 años (2015-2019)

- time
- generation biomass
- generation fossil brown coal/lignite
- generation fossil coal-derived gas
- generation fossil gas
- generation fossil hard coal
- generation fossil oil
- generation fossil oil shale

- generation fossil peat
- generation geothermal
- generation hydro pumped storage aggregated
- generation hydro pumped storage consumption
- generation hydro run-of-river and poundage
- generation hydro water reservoir
- generation marine
- generation nuclear
- generation other
- generation other renewable
- generation solar
- generation waste
- generation wind offshore
- generation wind onshore
- forecast solar day ahead
- forecast wind offshore eday ahead
- forecast wind onshore day ahead
- total load forecast
- total load actual
- price day ahead
- price actual

## 2. Exploración descriptiva del dataset

Hay algunas columnas que no poseen información, por lo que procederemos a eliminarlas directamente de la data set. También reemplazamos los valores faltantes por cero, para no tener problemas a la hora de entrenar los modelos.

| | generation biomass | generation fossil brown coal/lignite | generation fossil gas | generation fossil hard coal | generation fossil oil | generation hydro pumped storage consumption | generation hydro run-of-river and poundage | generation hydro water reservoir | generation nuclear | generation other | generation other renewable |
|---|---|---|---|---|---|---|---|---|---|---|---|
| count | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 |
| mean | 383.305727 | 447.829198 | 5619.851072 | 4253.880903 | 298.158139 | 475.319644 | 971.589351 | 2603.777407 | 6260.870123 | 60.197667 | 85.595739 |
| std | 85.796305 | 354.622834 | 2204.946787 | 1963.465684 | 52.963429 | 792.269198 | 401.307115 | 1835.677348 | 850.714243 | 20.279128 | 14.207003 |
| min | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 | 0.000000 |
| 25% | 333.000000 | 0.000000 | 4125.000000 | 2524.750000 | 263.000000 | 0.000000 | 637.000000 | 1076.000000 | 5757.000000 | 53.000000 | 73.000000 |
| 50% | 367.000000 | 509.000000 | 4968.000000 | 4473.000000 | 300.000000 | 67.000000 | 905.000000 | 2163.000000 | 6563.500000 | 57.000000 | 88.000000 |
| 75% | 433.000000 | 757.000000 | 6428.000000 | 5837.000000 | 330.000000 | 615.000000 | 1250.000000 | 3756.250000 | 7024.000000 | 80.000000 | 97.000000 |
| max | 592.000000 | 999.000000 | 20034.000000 | 8359.000000 | 449.000000 | 4523.000000 | 2000.000000 | 9728.000000 | 7117.000000 | 106.000000 | 119.000000 |

***Figura 1.*** *Dataset*

| generation solar | generation waste | generation wind onshore | forecast solar day ahead | forecast wind onshore day ahead | total load forecast | total load actual | price day ahead | price actual |
|---|---|---|---|---|---|---|---|---|
| 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 | 35064.000000 |
| 1431.930470 | 269.306126 | 5461.674595 | 1439.066735 | 5471.216689 | 28712.129962 | 28667.476928 | 49.874341 | 57.884023 |
| 1680.002043 | 50.572208 | 3215.250084 | 1677.703355 | 3176.312853 | 4594.100854 | 4664.083855 | 14.618900 | 14.204083 |
| 0.000000 | 0.000000 | 0.000000 | 0.000000 | 237.000000 | 18105.000000 | 0.000000 | 2.060000 | 9.330000 |
| 70.000000 | 240.000000 | 2930.750000 | 69.000000 | 2979.000000 | 24793.750000 | 24800.000000 | 41.490000 | 49.347500 |
| 615.000000 | 279.000000 | 4847.000000 | 576.000000 | 4855.000000 | 28906.000000 | 28894.000000 | 50.520000 | 58.020000 |
| 2575.250000 | 310.000000 | 7397.000000 | 2636.000000 | 7353.000000 | 32263.250000 | 32186.250000 | 60.530000 | 68.010000 |
| 5792.000000 | 357.000000 | 17436.000000 | 5836.000000 | 17430.000000 | 41390.000000 | 41015.000000 | 101.990000 | 116.800000 |

***Figura 2.*** *Dataset*

Finalmente, Después de la limpieza de datos, obtuvimos 20 columnas con los datos más relevantes para el estudio y entrenamiento.

Como nuestra variable de interés es el precio, queremos visualizar su comportamiento y para ello utilizamos un histograma de frecuencias, veremos cuál es el intervalo en el que más oscila.

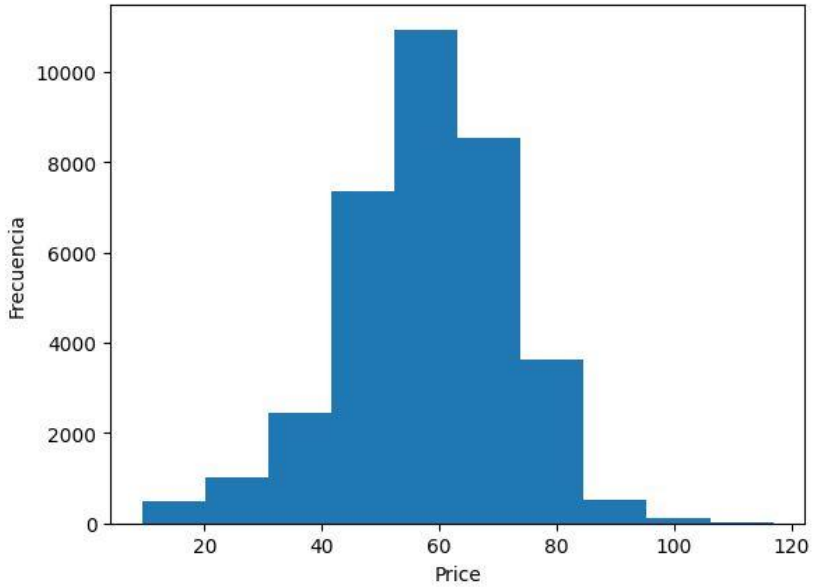


***Figura 3****. Histograma de Frecuencias del precio*

## 3. Iteraciones de desarrollo.

Para la primera iteración o modelo, tomamos el precio como una serie de tiempo, donde el precio vario cada hora como ya lo habíamos mencionado anteriormente. En este caso lo que se quiere hacer es predecir si el precio va a subir o va a bajar, volviendo así el trabajo en una tarea de clasificación, donde 1 es que aumenta y 0 que disminuye.

Para esto se usa el método del 80% entrenamiento y 20% test, por lo que los datos que se usarán para train y test serán así respectivamente

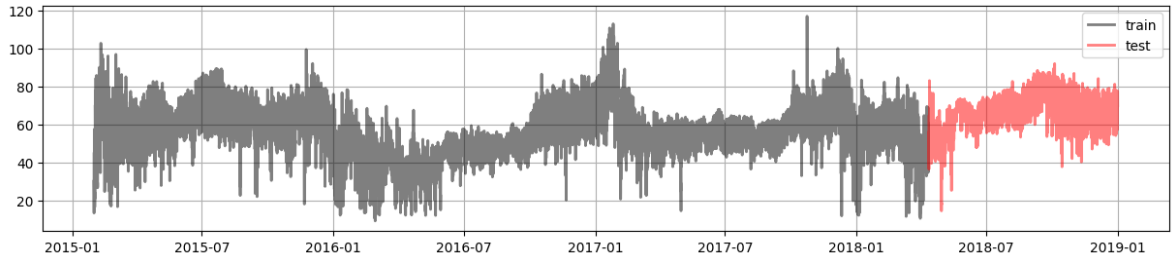


***Figura 4.*** *Fecha vs precio KWh*

Luego de tener claro los porcentajes de los datos, se construirá un data set con n datos atrás y el dato en la posición siguiente, como se muestra en la figura siguiente

| time | price actual_0 | price actual_1 | price actual_2 | price actual_3 | price actual_4 | price actual_5 | price actual_6 | price actual_7 | price actual_8 | price actual_9 | price actual_10 | price actual_11 | price actual_12 | price actual_13 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2015-01-31 00:00:00+01:00 | 65.41 | 64.92 | 64.48 | 59.32 | 56.04 | 53.63 | 51.73 | 51.43 | 48.98 | 54.20 | 58.94 | 59.86 | 60.12 | 62.05 |
| 2015-01-31 01:00:00+01:00 | 64.92 | 64.48 | 59.32 | 56.04 | 53.63 | 51.73 | 51.43 | 48.98 | 54.20 | 58.94 | 59.86 | 60.12 | 62.05 | 62.06 |
| 2015-01-31 02:00:00+01:00 | 64.48 | 59.32 | 56.04 | 53.63 | 51.73 | 51.43 | 48.98 | 54.20 | 58.94 | 59.86 | 60.12 | 62.05 | 62.06 | 59.76 |
| 2015-01-31 03:00:00+01:00 | 59.32 | 56.04 | 53.63 | 51.73 | 51.43 | 48.98 | 54.20 | 58.94 | 59.86 | 60.12 | 62.05 | 62.06 | 59.76 | 61.18 |
| 2015-01-31 04:00:00+01:00 | 56.04 | 53.63 | 51.73 | 51.43 | 48.98 | 54.20 | 58.94 | 59.86 | 60.12 | 62.05 | 62.06 | 59.76 | 61.18 | 64.74 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2018-12-31 19:00:00+01:00 | 72.72 | 70.48 | 69.72 | 67.69 | 64.50 | 60.27 | 53.37 | 51.32 | 50.03 | 50.25 | 50.98 | 51.73 | 51.42 | 52.01 |
| 2018-12-31 20:00:00+01:00 | 70.48 | 69.72 | 67.69 | 64.50 | 60.27 | 53.37 | 51.32 | 50.03 | 50.25 | 50.98 | 51.73 | 51.42 | 52.01 | 55.41 |
| 2018-12-31 21:00:00+01:00 | 69.72 | 67.69 | 64.50 | 60.27 | 53.37 | 51.32 | 50.03 | 50.25 | 50.98 | 51.73 | 51.42 | 52.01 | 55.41 | 60.80 |

***Figura 5.*** *Dataset*

Para este caso el n que se definió fue el de 720 datos, que corresponden a un mes en el historial de datos, acá los x serán los n valores que se tomen hasta llegar a la columna Price actual, y los y serán justamente la columna del Price actual.

Aquí se hace importante mencionar que el data frame que se muestra en la imagen se construye por una de las librerías que el profesor usa en uno de los notebooks del curso.

Posteriormente se crean los arrays, ya que el modelo no recibe series de tiempo, y por último se importa la librería, se emplea el modelo y se emplea una métrica para evaluar su desempeño. En este caso es una matrix confusión, donde se muestran los verdaderos positivos, los falsos positivos, los verdaderos negativos, y los falsos positivos, Con esto también se obtiene una un acurracy del modelo, es decir el porcentaje de acierto, que para el caso de la prueba es del 72%, que es verdaderamente aceptable.

A continuación, se muestra la matrix confusión

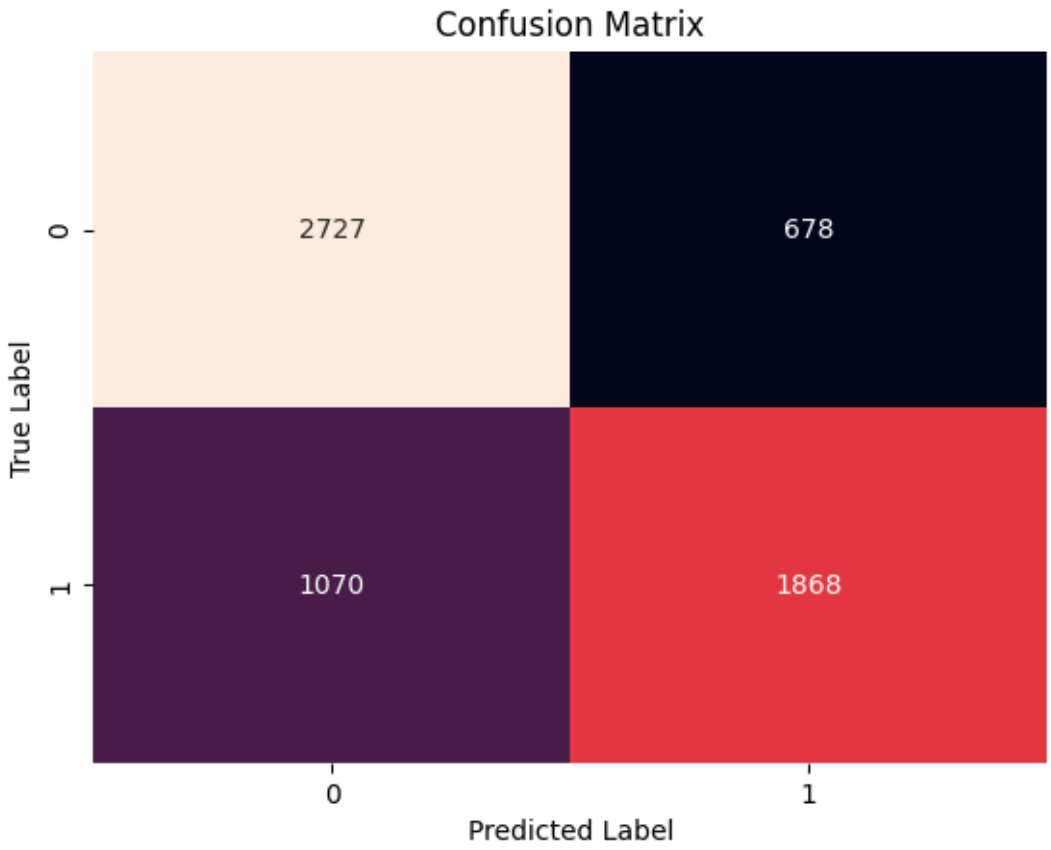


***Figura 6.*** *Matrix confusión.*

Para la segunda iteración, se quiere hacer una tarea de regresión multivariable, para predecir el precio, esta vez no si sube o baja sino un aproximado, por lo que escogimos los datos que más correlación tenían con este, estos se mostraran en la siguiente matriz de correlación.

Se puede observar que la relación es aproximadamente media, y por la distribución que muestran los datos del precio, se puede esperar desde un principio que la regresión se puede ajustar medianamente bien.

Realizamos nuevamente la separación de los datos en x train, x test, y train e y test, donde las x serán las variables elegidas mediante la matriz de correlación e y será nuevamente Price actual. También se vuelve a mencionar que se usa la regla del 80% train y 20% test.

Se hace el procedimiento correspondiente para dejar los datos a usar en la forma de array que se usa para alimentar el modelo.

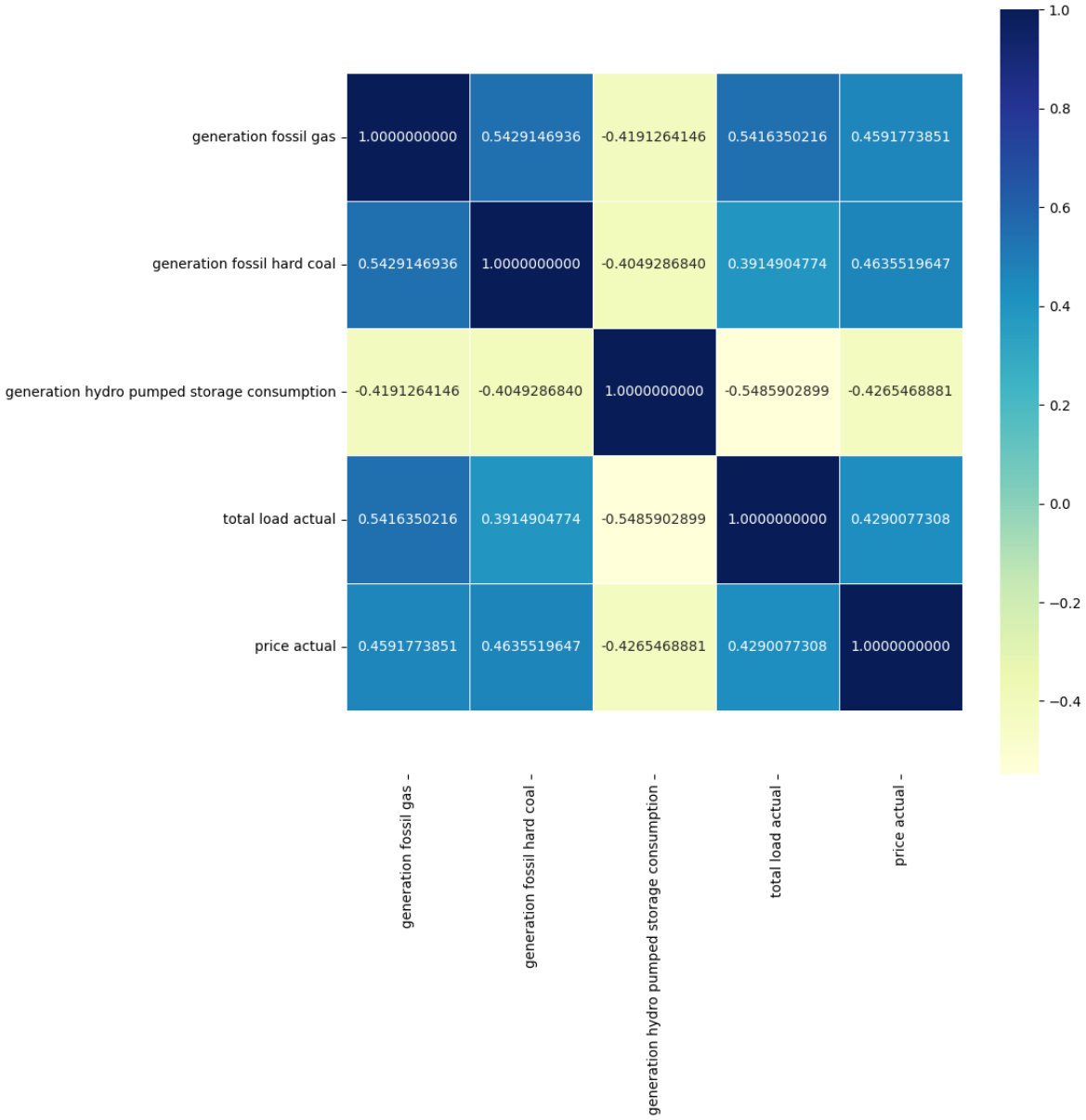


***Figura 7.*** *Matriz de correlación.*

Luego de importar las librerías, se emplea el modelo y procedemos a graficar los datos de la prueba junto con los datos predichos, para ver cómo se comportan y que tan cercanos o lejanos están, esto dará una idea visual de como se comporta el modelo.

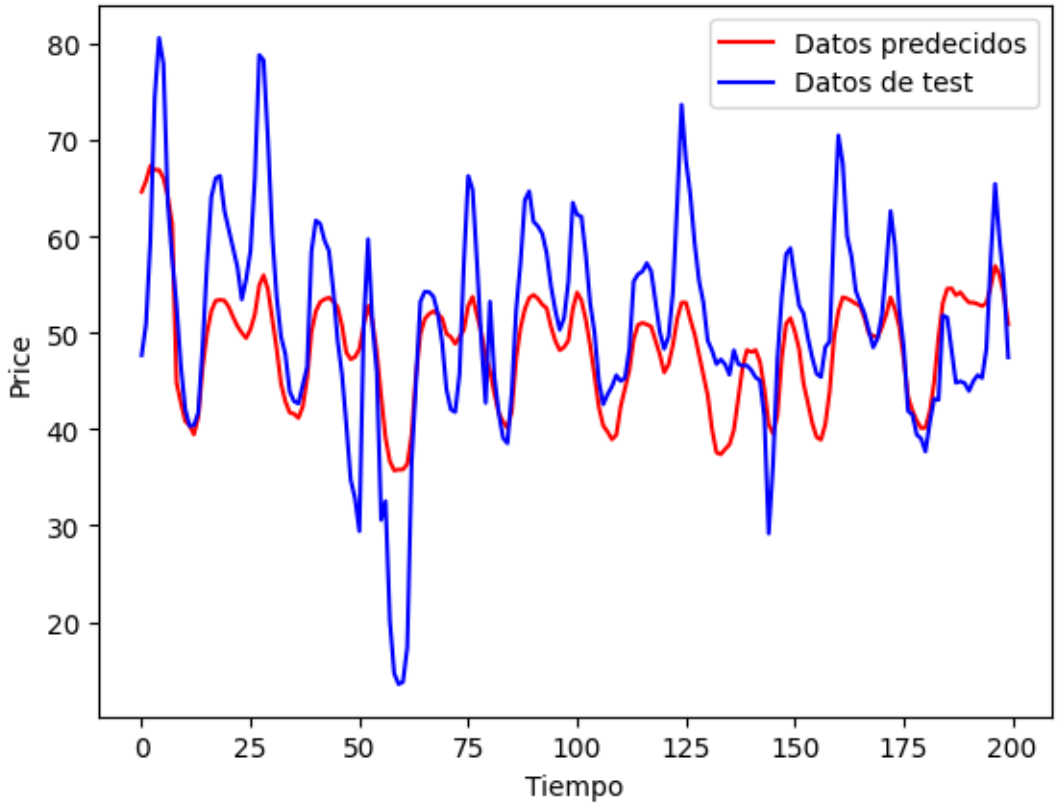


***Figura 8.*** *Tiempo vs Precio KWh*

Posteriormente y para medir de una manera más estricta el desempeño del modelo, se usa la medida de error RSMLE, la cual es 0.23, lo que dice que el modelo empleado tiene una confiablidad de aproximadamente el 70%.

Con lo anterior podemos decir que los dos modelos empleados funcionan bien, y hay que ser claros en que sirven para cosas diferentes.

## 4. Retos y consideraciones de despliegue.

Por ser estudiantes de ingeniería eléctrica, queríamos orientar nuestro trabajo a modelos que se pueden usar en el sector, y justamente la demanda, la oferta y el precio que en muchos países se tranza en bolsa, también en un principio se quiso hacer con datos del país, pero la disponibilidad de los datos y la desorganización de los mismos hicieron que fuera difícil seguir ese camino, ya que se entiende que la finalidad del curso no es estrictamente la minería de datos, por eso se opto por un data set del país de España, donde se contaba con más datos y en un mejor orden.

Como consideraciones generales, es importante tener una idea clara de lo que se quiere hacer y con que modelo se quiere, para poder seleccionar de buena manera los datos con los que este se va a entrenar, pues en este caso si usáramos una regresión con los datos del primer caso que era de clasificación, seguramente se obtendrían resultados no deseados o que no sirven de nada, lo mismo pasaría si el algoritmo de clasificación se entrena con las variables usadas en la regresión multivariable. Por eso es importante tener una idea medianamente de cómo funciona.

En el caso de que se fuera a vender el modelo a un cliente, es importante entender para qué se le va a dar uso a este, pues ya que, si un cliente desea tener una predicción a largo plazo, lo ideal no sería venderle un modelo de clasificación, ya que la mayor utilidad que este tiene podemos decir que es en intervalos pequeños de tiempo, según su necesidad.

## 5. Conclusiones

-Hay aplicaciones de modelos de machine learning que no nos imaginamos, pero que de alguna manera se han vuelto parte de la vida cotidiana y no lo sabemos, pues un modelo como estos ayuda a tener estabilidad en el precio de un bien tan necesario como lo es la electricidad.

-Este tipo de modelos permiten simular escenarios probables, como un fenómeno del niño, o un escenario geopolítico que altere los precios y la disponibilidad de estos recursos, con esto una comisión de regulación puede tomar decisiones de prevención para asegurar su abastecimiento de energía eléctrica.

-Gracias a estos modelos predictivos se puede determinar que el precio probablemente seguirá creciendo, ya que, si se hace la comparación con los datos reales y los predichos se tiene una aproximación cercana, estas predicciones podrían mejorar si se tienen en cuenta los recursos para la generación y lo escasos que se están volviendo poco a poco.

**Bibliografía**


https://www.kaggle.com/code/nathanoliver/spain-electricity-price-prediction/notebook

https://www.kaggle.com/code/dimitriosroussis/electricity-price-forecasting-with-dnns-eda/notebook#1.-Exploration-and-Cleaning

https://platzi.com/tutoriales/1794-pandas/6926-usando-la-api-de-kaggle-con-google-colab-para-carga-y-descarga-de-datasets/