



**Business  
School**

MÁSTER EN DATA SCIENCE Y BUSINESS ANALYTICS  
ONLINE

**ANTEPROYECTO PARA EL TFM  
“PREDICCIÓN DEL PRECIO DE LA  
ELECTRICIDAD EN EL MERCADO  
ESPAÑOL”**

ANTEPROYECTO elaborado por: **Iñigo Elorza Barea**  
Tutor/a de TFM: **Abel Ángel Soriano Vázquez**

- Madrid a 24 de abril de 2023 -

## 1. TÍTULO

El título dado al Proyecto es: **“Predicción del precio de la electricidad en el mercado español”**.

## 2. CONTEXTO DE NEGOCIO O TECNOLÓGICO

El sector eléctrico en España es un mercado altamente complejo debido a los numerosos factores que le afectan. Entre estos están:

- El **equilibrio que debe existir entre la generación y la demanda** en cada instante.
- Las **fuentes de producción de energía eléctrica es muy diversa** (hidráulica, eólica, nuclear, solar, térmica, etc).
- **Cada tecnología de producción tiene un coste de producción** asociado, ya sea debido a las infraestructuras necesarias o para la propia producción de la energía.
- También existe un **coste ambiental** como es la cantidad de **emisiones de CO<sub>2</sub>**.
- Las fuentes de **energía renovables dependen de las condiciones meteorológicas**.
- La **legislación del sector** también aumenta la complejidad en el sector.

Todo esto hace que **la predicción del precio de la electricidad se convierta en un elemento esencial en la toma de decisiones** en las compañías del sector, no sólo para desarrollar estrategias de generación para las Productoras, sino también para otros actores del mercado como Comercializadoras, que adquieren energía para su venta al Consumidor final o para los Consumidores Directos en Mercado, que son grandes consumidores (industria).

Gracias la capacidad computacional actual y al desarrollo de software capaz de manejar con efectividad grandes volúmenes de datos, se pueden emplear herramientas **de Deep Learnin**, y especialmente las **redes neuronales, para predecir el precio de la electricidad a largo plazo**.

## 3. OBJETIVO DEL PROYECTO

En la documentación consultada hasta el momento se han encontrado pocas referencias a estudios para la predicción de precios de la electricidad a más de tres meses. Por este motivo, **la predicción del precio para el mercado a plazo es un territorio muy desaprovechado** y supone una oportunidad para realizar un estudio con el **objetivo de aportar nuevos mecanismos en la toma de decisiones** para las compañías del sector.

El **objeto del Trabajo de Fin de Máster** es **desarrollar una herramienta de predicción de precios de la electricidad a plazo en el mercado español**.

Se pretende realizar partiendo de una serie de **variables**, como son los **precios la las materias primas que se emplean para la producción de electricidad** (gasóleo, gas natural y carbón), **producción de cada por tipo de central, demanda eléctrica, meteorología** (temperatura, viento y agua embalsamada) y otros factores (**derechos de emisión de CO<sub>2</sub> y situación económica del país**) y empleando técnicas de **análisis de datos**, técnicas de **aprendizaje automático** como son las series temporales y técnicas de **Deep Learning** (redes neuronales).

## 4. RESULTADOS ESPERADOS

Del artículo de investigación realizado por González C., Mira-McWilliams J. y Juárez I. (2015). Evaluación de variables importantes y pronóstico del precio de la electricidad basado en modelos de árboles de regresión: clasificación y árboles de regresión, Bagging y Random Forests. IET Generation, Transmission & Distribution, 9(11), 1120–1128, se pueden extraer los siguientes **errores medios en las predicciones de los precios de la electricidad según el método empleado en distintos estudios e investigaciones realizados**.

FUENTE	METODO o MODELO/S EMPLEADO/S	TIPO DE MERCADO	DATOS EMPLEADOS	ERROR MEDIO DE PREDICCION
Mori, H., Awata, A.: 'Extracción de datos de pronóstico de precios de electricidad con árbol de regresión y red de función de base radial normalizada'. proc. del IEEE Int. Conf. sobre Sistemas, Hombre y Cibernética, 2007, ISIC, (doi:10.1109/ICSMC.2007.4414228)	Método híbrido de árboles de regresión y red de función de base radial normalizada	A corto plazo (con una hora de antelación)	Mercado de Nueva Inglaterra del 1 al 31 de julio de 2005 y del 1 al 7 de agosto de 2005	10,36%
Neupane, B., Perera, KS, Aung, Z., Woon, WL: 'Pronóstico de precios de electricidad basado en redes neuronales artificiales para el despliegue de redes inteligentes'. En t. Conf. sobre Sistemas Informáticos e Informática Industrial (ICCSII), del 18 al 20 de diciembre de 2012 (ISBN: 978-1-4673-5155-3. doi: 10.119/ICCSII.2012.6454392)	Red neuronal	A corto plazo (con una hora de antelación)	Mercado español de 2004 y 2005 para entrenamiento y de 2006 para pruebas	5,76%
Troncoso, A., Riquelme, JM, Gómez, A., Martínez, JL, Riquelme, JC: 'Pronóstico de precios de mercado de electricidad basado en técnicas de vecinos más cercanos ponderados', Trans. IEEE. Sistema de energía, 2007, 22, págs. 1294–1301	Técnicas ponderadas de vecinos más cercanos	Diario (con un día de antelación)	Mercado español 2002	9,80%
Catalão, JPS, Pousinho, HM, Mendes, VMF: 'Precios de electricidad a corto plazo: previsión en un mercado competitivo mediante un enfoque híbrido inteligente', Convertidores de energía. Gerente, 2011, 52, págs. 1061–1065	Redes neuronales y lógica difusa	A plazo (con una semana de antelación)	Mercado español 2002	9,44%
Alonso, A., García-Martos, C., Rodríguez, J., Sánchez, MJ: 'Análisis factorial dinámico estacional e inferencia bootstrap: aplicación a la previsión del mercado eléctrico', tecnometría, 2011, 53, págs. 137–151	Análisis factorial dinámico estacional	Diario (con un día de antelación)	Mercado español 2004	11,00%
García-Martos, C., Rodríguez, J., Sánchez, MJ: 'Predicción de los precios de la electricidad mediante la extracción de factores comunes dinámicos: aplicación al Mercado Ibérico', IET General. Transm. Distrib., 2011, 1, págs. 1 a 10	Modelo de factor dinámico	Diario (con un día de antelación)	Mercado español 2007-2009	7,39%
García-Martos, C., Rodríguez, J., Sánchez, MJ: 'Predicción de los precios de la electricidad y sus volatilidades utilizando componentes no observados', Economía de energía, 2011, 33, págs. 1227–1239	Análisis factorial dinámico estacional heterocedástico	Diario (con un día de antelación)	Mercado español 2008	5,76%

Observando estos errores medios obtenidos en distintos métodos y modelos predictivos, **podríamos considerar que un buen resultado para los modelos estudiados en el TFM sería un error medio por debajo del 11%.**

## 5. ASIGNATURAS O MÓDULOS RELACIONADOS CON LOS OBJETIVOS Y RESULTADOS

El TFM se basará principalmente en los siguientes módulos impartidos en el Master:

- **MODULO I: Las Herramientas del Científico de Datos.**
- **MODULO II Impacto y valor del Big Data.**
- **MODULO III La Ciencia de Datos. Técnicas de análisis, minería y visualización.**
- **MODULO V Estadística para Científicos de Datos.**
- **MODULO VII: Aprendizaje automático.**
- **MODULO VIII: Inteligencia artificial para la empresa.**

Como es lógico e ineludible, para desarrollar el TFM, hay que emplear los conocimientos adquiridos en el **MODULO I**, ya que es donde se desarrollan los conocimientos sobre los dos lenguajes de programación (**Python** y **R**) a través de los cuales se realizará la recolección, la limpieza y el análisis de los datos y se desarrollarán los modelos de predicción. **La idea inicial es emplear los dos lenguajes de programación**, aunque esto podría variar según vaya avanzando el proyecto debido a

impedimentos técnicos o a encontrarnos con altas dificultades para resolver algún aspecto determinado con el lenguaje previsto inicialmente.

Tanto el **MODULO II Impacto y valor del Big Data**, como el **MODULO III La Ciencia de Datos. Técnicas de análisis, minería y visualización**, son básicos e imprescindibles para poder entender los módulos posteriores y para poseer los conceptos necesarios para poder actuar correctamente con los datos (obtención, limpieza, transformación y visualización).

El **MODULO V Estadística para Científicos de Datos** es básico para la comprensión del **análisis estadístico de los datos en el que se basan los algoritmos de Machine Learning**.

El **MODULO VI** es necesario ya que en él se exponen y desarrollan dos temáticas fundamentales en el TFM como son, los **algoritmos de Aprendizaje Automático**, más concretamente los empleados para resolver los **problemas de regresión**, y las técnicas de **Deep Learning**, en concreto las **redes neuronales**. Estas dos técnicas serán las empleadas **para realizar la previsión de los precios**.

Finalmente, el **MODULO VII** será también necesario porque en él se abordan las **técnicas para la toma de decisiones**, y más concretamente lo referente al **Aprendizaje Supervisado** en el que se tratan los **algoritmos para problemas de regresión** y las **redes neuronales**.

## **6. DATASETS QUE SE VAN A EMPLEAR Y MANERA DE SU OBTENCION, VIABILIDAD DEL PROYECTO**

Los datasets empleados serán obtenidos de diferentes fuentes a través de internet. Estas son:

- Precio de la electricidad: <https://www.ree.es> (Red Eléctrica de España).
- Producción de los diferentes tipos de centrales: <https://www.ree.es> (Red Eléctrica de España).
- Demanda de electricidad en España: <https://www.ree.es> (Red Eléctrica de España).
- Precio de las materias primas: <https://es.investing.com> (plataforma de mercados financieros).
- Datos meteorológicos: <https://www.aemet.es/> (Aemet).
- Agua embalsada: <https://www.epdata.es/> (Plataforma para facilitar el uso de datos públicos).
- Derechos de emisión de CO2 y situación económica del país <https://es.investing.com> (plataforma de mercados financieros).