

PROYECTO DS



Camino hacia la transición energética - Una perspectiva desde el mercado eléctrico

Segunda Entrega del Proyecto Final - Obtención de Insights a partir de visualizaciones -



Comunidad Coderhouse

Curso Data Science

Plantel docente: Prof. Ruiz, Jorge & Tutora Ruscitti, Aldana

Alumno: Risculese Francisco

Abstracto



Abstracto

La estimación del precio de la energía eléctrica es un desafío crucial en el contexto de la sostenibilidad y la transición energética, temas de gran relevancia en la actualidad. Para abordar este desafío, es fundamental comprender en detalle el comportamiento de la variable a estimar, identificar a los actores clave en el mercado energético y analizar las complejidades que influyen en este sector.

El incremento en los precios de los combustibles fósiles, así como su dependencia en muchos países, ha resultado en un aumento significativo en el precio de la energía eléctrica. Sin embargo, esta situación ha impulsado el interés en soluciones energéticas renovables a corto plazo. En España, el gobierno está acelerando la aprobación de proyectos que se centran en la sostenibilidad y la protección del medio ambiente. Esto incluye la adopción de tecnologías avanzadas y promueve aspectos positivos como la autogestión y el autoconsumo de energía.

Objetivo



Objetivo

Predicción del precio de la energía eléctrica en España, mediante modelos de regresión, para el período 2015-2018 empleando un dataset obtenido del website "Kaggle".

Contexto comercial



Contexto comercial

Se debe conocer de antemano que el aumento de los precios de los combustibles fósiles, como el gas natural y el carbón, ha provocado un aumento de los precios de la energía eléctrica. Este aumento es especialmente notable en los países que dependen en gran medida de los combustibles fósiles para generar electricidad.

Las energías renovables (hidroeléctricas, solar o eólica) son una fuente de energía más barata y fiable que los combustibles fósiles.

Además, su aplicación y campo de acción ha ocupado un buen espacio en el sector de la energía para electrificación, según datos de consumo medidos en el año 2019.

Renewable Energy in Total Final Energy Consumption, by Final Energy Use, 2019



Note: Data should not be compared with previous years because of revisions due to improved or adjusted methodology.

REN21 RENEWABLES 2022 GLOBAL STATUS REPORT

Source: Based on IEA data.

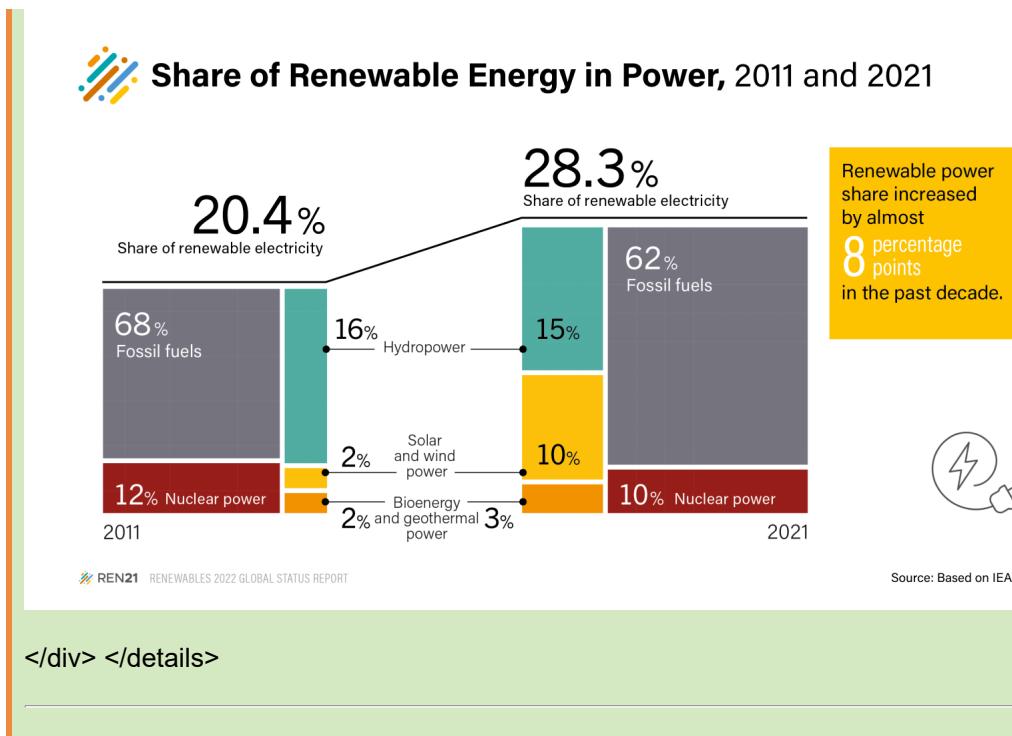
En materia de costos, las energías solar y eólica han disminuido significativamente su valor en los últimos años, lo que las ha hecho más competitivas con respecto a las fuentes de energía convencionales (combustibles fósiles). El aumento de la participación de estas energías limpias en la matriz energética apoya a la reducción de los precios de la energía eléctrica.

La determinación del precio sigue una secuencia de actividades que involucra a todos los actores de la cadena de negocio (explicado a continuación), pero para entenderlo desde una perspectiva clara desde el precio: se debe conocer que aquellas energías poco flexibles que ofertan su producción a costos bajos (solar y eólica) o aquellas que ofertan su producción de forma constante y a plena potencia, como la nuclear, podrán disminuir con sus ofertas el precio de este servicio. Esto se justifica porque las tecnologías mencionadas ofertan de manera necesaria toda la energía que producen, ya que no poseen la capacidad de almacenar y gestionar sus recursos (son, como se dijo, poco flexibles).

Por otro lado, aquellas que ofertan su generación con costos de producción elevados (las que emplean combustibles fósiles), hacen que el precio suba, y pueden participar con mucha más flexibilidad frente a los requerimientos de la demanda. Un comportamiento similar ocurre con las centrales hidroeléctricas que pueden gestionar de forma práctica el recurso hídrico, pero con la diferencia que poseen un costo de obtención del recurso nulo, al ser obtenido del medio ambiente.

En España, el precio de la electricidad se ha multiplicado por cuatro en los últimos años. Este aumento es debido a la creciente dependencia de los combustibles fósiles para generar electricidad. Por ello, el gobierno español está acelerando la transición a las energías renovables. Fuentes gubernamentales proponen aumentar la participación de las energías renovables en la matriz energética española del 40% al 74% para 2030.

Como España, muchos países se involucran en este movimiento acelerado y los cambios en un plazo de 10 años, a nivel mundial, están tomando cada vez más impronta, sobre todo en el sector de la electrificación.



</div> </details>

Problema comercial

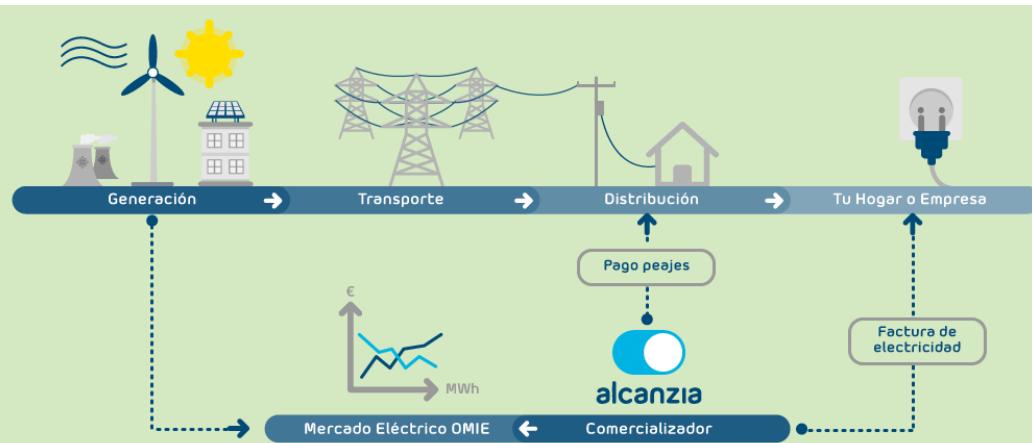


Problema comercial

Como científico de datos, se desempeñará la labor en estimar el precio de la energía eléctrica para la empresa OMIE. Es obligatorio conocer el comportamiento del mercado eléctrico español, y el de sus actores participantes. Siendo el OMIE el operador del Mercado Ibérico de Energía, que opera tanto en Portugal como en España: este tiene diversas funciones en todo el proceso de intercambio de energía, regulando el proceso e interviniendo en ciertas circunstancias, por ejemplo en acuerdos entre los comercializadores (compradores de energía) y los generadores (vendedores de energía).

En forma resumida, participa como un actor terminal el sector de los generadores, que son los destinados a producir la energía eléctrica, sea generada por fuentes convencionales de energía (como la nuclear o los combustibles fósiles) o por fuentes renovables (destacándose la hidroeléctrica y la eólica).

Luego, la matriz de transportistas, distribuidores y comercializadores, por separado, brindarán las condiciones necesarias para que el servicio sea trasladado, por medio de las redes de alta-media-baja tensión, al otro actor terminal que forma parte del negocio: el sector consumidor.



A partir de ciertas evaluaciones, se buscará hacer uso del mejor modelo que se adapte a cumplir con el objetivo y afrontar el problema comercial, guiado por una serie de interrogantes e hipótesis planteadas, las cuales se desarrollarán en el EDA (Exploratory Data Analysis).

Preguntas principales

- ⚡ Variable precio. ¿Qué resultados se observan, haciendo un análisis estadístico descriptivo e inferencial, sobre la variable target?
- ⚡ Evolución temporal para el precio. ¿Cómo se comporta el precio de la energía eléctrica a lo largo del período considerado? ¿Cómo se lleva a cabo la descomposición de la serie temporal, y qué se puede determinar realizando ésto? ¿Qué sucede con la estacionalidad de la serie temporal?
- ⚡ Generación energética y precio de la electricidad. ¿Cómo se da la evolución en el aporte de cada fuente de generación energética en un período anual? ¿Y cómo se da la evolución en el aporte de cada fuente de generación energética en un período diario? ¿Qué relación se puede obtener si se incluye la variable precio a estas visualizaciones?
- ⚡ Valores atípicos de la variable precio. ¿Qué análisis se puede realizar teniendo en cuenta los valores outliers?
- ⚡ Generación de energía por ubicación geográfica. Si el conjunto de datos incluye información sobre la ubicación de las plantas de generación de energía, ¿cómo varía la generación de energía en diferentes regiones geográficas? ¿Qué se obtiene de un análisis multivariado?

Preguntas complementarias

- ⚡ Evolución de la demanda de energía en el país. ¿Cómo evoluciona la situación de esta variable para los años de estudio? ¿Qué comportamientos se muestran con respecto a su tendencia? ¿Qué se puede concluir respecto a un período semanal de consumo?
- ⚡ Evolución temporal para el precio. ¿Cómo se comporta el precio de la energía eléctrica analizando cada año en particular?
- ⚡ ¿Qué relación se puede establecer, a grandes rasgos, entre la variable demanda y el precio fijado de la energía eléctrica?
- ⚡ Relaciones con la variable precio. ¿Qué relación se puede establecer, a grandes rasgos, entre la variable target y las demás features?
- ⚡ Generación energética. ¿Qué se puede observar cuando se analiza la distribución de valores en un histograma para cada variable?
- ⚡ Valores atípicos de la variable precio. ¿Qué comportamiento presenta la variable target, analizando sus valores atípicos?
- ⚡ Valores atípicos de las variables climatológicas. ¿Qué comportamientos presentan las variables, analizando sus valores atípicos? ¿Qué toma de decisión se puede realizar con el análisis?
- ⚡ Generación energética. ¿Cuál es el aporte absoluto de cada fuente de energía (renovable o convencional) a su generación total? ¿Qué comportamiento exhibe cada fuente?
- ⚡ Proporción en la generación. ¿Cuál es la proporción porcentual de generación de energía proveniente tanto de fuentes renovables como de fuentes convencionales?
- ⚡ Impacto ambiental. ¿Existe alguna correlación entre la generación de energía a partir de fuentes fósiles y las emisiones de gases de efecto invernadero u otros impactos ambientales?
- ⚡ Potencia instalada en el sector de generación energética. ¿Cómo se distribuye la cantidad de potencia instalada para las diversas tecnologías de producción de energía eléctrica? </div> </details>

Contexto analítico



Contexto analítico

Se cuentan con dos datasets, bajo un nombre genérico que identifica a ambos. Estos están en formato .csv.

Uno se utilizó para desarrollar los primeros 6 desafíos entregables y el restante contiene información adicional y complementaria que se utilizó para desarrollar los restantes 6 desafíos entregables, que contenían un complemento a la Data Acquisition mediante APIs públicas, Data Wrangling y Data Storytelling.

Dataset general: Hourly energy demand generation and weather. El dataset general contiene, en un período de 4 años, datos sobre el consumo, generación, precio y climatología en España.

energy_dataset.csv: Contiene datos sobre consumo y generación, recopilados desde el portal público del ENTSOE-E (European Network of Transmission System Operators) para los operadores de red de transporte TSO (Transmission Service Operator). La declaración de los precios fue obtenida por el TSO de España: Red Eléctrica España.

weather_features.csv: La información climatológica fue comprada como parte de un proyecto personal, desde el [Open Weather API](#), para las 5 ciudades más grandes del país, y empleadas para este proyecto (Madrid, Barcelona, Valencia, Sevilla y Bilbao).

También se emplea información adicional:

- Portal público del ENTSOE-E (European Network of Transmission System Operators)
- Website del OMIE (ente regulador del precio)
- Website de la REE (Red Eléctrica de España) - ente del transporte energético -
- Centro de descargas del CNIG
- Informes anuales de la REE
- Bibliografía consultada
- Valor atípico - Wikipedia

Se puede realizar un breve análisis descriptivo de las variables que contienen ambos datasets, detallando sus unidades de medida y colocando una breve descripción con respecto al recurso energético que figura. Todas las medidas son para el país de España, a nivel nacional.

Primer dataset

- 'time': Es el tiempo medido desde el 31 de diciembre del 2014 hasta el 31 de diciembre del 2018, en escala diaria y en intervalos de 1 hora. Siguiendo el estándar del [CET \(Central European Time\)](#).
- 'generation biomass': Es la generación eléctrica a partir del recurso renovable [biomasa](#) - Unidades en MW -
- 'generation fossil brown coal/lignite': Es la generación eléctrica a partir del combustible fósil [Lignito/carbón marrón](#) - Unidades en MW -
- 'generation fossil coal-derived gas': Es la generación eléctrica a partir del combustible fósil [gas de carbón](#) - Unidades en MW -
- 'generation fossil gas': Es la generación eléctrica a partir del [gas fósil natural](#) - Unidades en MW -
- 'generation fossil hard coal': Es la generación eléctrica a partir del [carbón fósil duro o Antracita](#) - Unidades en MW -
- 'generation fossil oil': Es la generación eléctrica a partir de [fueloil](#) - Unidades en MW -
- 'generation fossil oil shale': Es la generación eléctrica a partir de [Esquistos bituminosos fósiles \(Oil shale\)](#) - Unidades en MW -
- 'generation fossil peat': Es la generación eléctrica a partir del material fósil [Turba](#) - Unidades en MW -
- 'generation geothermal': Es la generación eléctrica renovable a partir de la [energía geotérmica](#) - Unidades en MW -
- 'generation hydro pumped storage aggregated': Es la generación eléctrica renovable a partir del [recurso hidráulico](#), considerando un almacenamiento mediante un sistema de bombeo añadido - Unidades en MW -
- 'generation hydro pumped storage consumption': Es la generación eléctrica renovable a partir del [recurso hidráulico](#), considerando un almacenamiento mediante un sistema de bombeo añadido - Unidades en MW -
- 'generation hydro run-of-river and poundage': Es la generación eléctrica renovable a partir del [recurso hidráulico](#), considerando una central hidroeléctrica fluyente/de pasada - Unidades en MW -
- 'generation hydro water reservoir': Es la generación eléctrica renovable a partir del [recurso hidráulico](#), considerando un reservorio/embalse - Unidades en MW -
- 'generation marine': Es la generación eléctrica renovable a partir del [recurso del mar](#) - Unidades en MW -
- 'generation nuclear': Es la generación eléctrica a partir de la [energía nuclear](#) - Unidades en MW -
- 'generation other': Es la generación eléctrica considerando otras fuentes energéticas, no consideradas en este listado de variables - Unidades en MW -
- 'generation other renewable': Es la generación eléctrica renovable considerando otras fuentes energéticas, no consideradas en este listado de variables - Unidades en MW -
- 'generation solar': Es la generación eléctrica renovable empleando la [energía solar](#) - Unidades en MW -
- 'generation waste': Es la generación eléctrica a partir del aprovechamiento energético de los [residuos](#) - Unidades en MW -
- 'generation wind offshore': Es la generación eléctrica renovable a partir del [recurso eólico marino, u "offshore"](#) - Unidades en MW -

- 'generation wind onshore': Es la generación eléctrica renovable a partir del [recurso eólico terrestre, u "onshore"](#) - Unidades en MW -
- 'forecast solar day ahead': Es la previsión del día siguiente al considerado, sobre la generación eléctrica renovable empleando la energía solar - Unidades en MW
- 'forecast wind offshore eday ahead': Es la previsión del día siguiente al considerado, sobre la generación eléctrica renovable a partir del recurso eólico marino, u "offshore" - Unidades en MW
- 'forecast wind onshore day ahead': Es la previsión del día siguiente al considerado, sobre la generación eléctrica renovable a partir del recurso eólico terrestre, u "onshore" - Unidades en MW
- 'total load forecast': Es la previsión del día siguiente al considerado, sobre la demanda eléctrica que tiene el país - Unidades en MW
- 'total load actual': Es la demanda eléctrica que tiene el país - Unidades en MW
- 'price day ahead': Es la previsión del día siguiente al considerado, sobre el precio de la energía eléctrica que tiene el país - Unidades en €/MWh
- 'price actual': Es el precio de la energía eléctrica que tiene el país - Unidades en €/MWh

Segundo dataset - Data Enrichment -

- 'dt_iso': Es el tiempo medido desde el 31 de diciembre del 2014 hasta el 31 de diciembre del 2018, en escala diaria y en intervalos de 1 hora. Siguiendo el estándar del CET (Central European Time).
- 'city_name': Alude al nombre de la ciudad para la cual se obtienen los datos de cada variable (incluyen las ciudades de Madrid, Barcelona, Sevilla, Valencia y Bilbao).
- 'temp': Valores medidos de temperatura, en grados Kelvin (°K).
- 'temp_min': Valores medidos para los mínimos de temperatura, para dicho momento temporal, en grados Kelvin (°K).
- 'temp_max': Valores medidos para los máximos de temperatura, para dicho momento temporal, en grados Kelvin (°K).
- 'pressure': Valores medidos para la presión atmosférica (a nivel del mar), en HPa - 1 HPa = 0,00099 atm = 0.0145 psi
- 'humidity': Valores medidos para la humedad, en porcentajes.
- 'wind_speed': Valores medidos para la velocidad del viento, en m/s.
- 'wind_deg': Valores medidos para la dirección del viento, en grados (meteorológicos).
- 'rain_1h': Volumen de lluvia para la última hora considerada, en mm.
- 'rain_3h': Volumen de lluvia para las últimas tres horas consideradas, en mm.
- 'snow_3h': Volumen de nieve para las últimas tres horas consideradas, en mm. y considerando el estado líquido.
- 'clouds_all': Porcentaje de nubosidad.
- 'weather_id': ID referido a la condición climática.
- 'weather_main': Grupo categórico de parámetros climáticos (lluvia, nieve, evento extremo, etc.)
- 'weather_description': Condición climática dentro del grupo categórico.
- 'weather_icon': Ícono climático asociado al ID.

Se debe complementar la información brindada de las últimas 4 variables con el contenido del siguiente [enlace](#). Toda la información, como se aclaró antes, proviene de [OpenWeather](#).

El siguiente [documento en formato .xlsx](#) hace una revisión de algunos parámetros importantes, referidos a las 5 principales Comunidades Autónomas (CC.AA.) del país ibérico. Se busca que a la hora de tratar y analizar los datos en esta notebook ya se cuente con una base de información estadística sobre parámetros como población, infraestructura, generación eléctrica y consumo eléctrico.

</div> </details>

Importación de librerías



Importación de Librerías

- Se importan las librerías correspondientes para efectuar los pasos del Proyecto de DS

```
In [1]: import pandas as pd
import numpy as np
```

```
import matplotlib as mpl
import matplotlib.pyplot as plt
import matplotlib.colors as mcolors
import matplotlib.dates as mdates
import seaborn as sns
```

```

import plotly.express as px
import plotly.graph_objects as go
import plotly.io as pio
import bokeh.plotting as bkplt
import graphviz
from bokeh.plotting import figure, show
from bokeh.models import ColorBar, ColumnDataSource, HoverTool
from bokeh.transform import linear_cmap
from bokeh.io import output_notebook
from bokeh.palettes import Viridis256

import pickle

import statsmodels.api as sm
import scipy.stats as stats
import scipy.stats.mstats as mstats
from statsmodels.tsa.stattools import adfuller
from statsmodels.graphics.tsaplots import plot_acf, plot_pacf
from sklearn.preprocessing import StandardScaler
from sklearn.feature_selection import f_classif
from sklearn.feature_selection import VarianceThreshold
from sklearn.feature_selection import f_regression, SelectKBest
from sklearn.metrics import mean_squared_error as mse
from sklearn.pipeline import Pipeline
from scipy.stats import pearsonr
from scipy.stats import shapiro
from sklearn.preprocessing import PolynomialFeatures
from sklearn.linear_model import LinearRegression
from sklearn.model_selection import train_test_split
from sklearn.metrics import mean_squared_error, r2_score, mean_absolute_error
from sklearn.neighbors import KNeighborsRegressor
from sklearn.tree import DecisionTreeRegressor
from sklearn.tree import plot_tree
from sklearn import tree

import warnings
warnings.filterwarnings("ignore", category=UserWarning)

```

Data Acquisition



Data Acquisition

```

In [2]: from google.colab import drive
import os
drive.mount ('/content/gdrive')

# 1º dataset - Hourly energy demand generation and weather
%cd '/content/gdrive/MyDrive'

# 1º dataset (energy_dataset.csv)
df_energy = pd.read_csv('Portfolio DS&A/Proyecto final - Data Science/PROYECTO DS - Risculese Francisco/1º Entrega

df_energy

Mounted at /content/gdrive
/content/gdrive/MyDrive

```

PROYECTO

Out[2]:

	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	ge
0	2015-01-01 00:00:00+01:00	447.0	329.0	0.0	4844.0	4821.0	162.0	0.0	0.0	0.0
1	2015-01-01 01:00:00+01:00	449.0	328.0	0.0	5196.0	4755.0	158.0	0.0	0.0	0.0
2	2015-01-01 02:00:00+01:00	448.0	323.0	0.0	4857.0	4581.0	157.0	0.0	0.0	0.0
3	2015-01-01 03:00:00+01:00	438.0	254.0	0.0	4314.0	4131.0	160.0	0.0	0.0	0.0
4	2015-01-01 04:00:00+01:00	428.0	187.0	0.0	4130.0	3840.0	156.0	0.0	0.0	0.0
...
35059	2018-12-31 19:00:00+01:00	297.0	0.0	0.0	7634.0	2628.0	178.0	0.0	0.0	0.0
35060	2018-12-31 20:00:00+01:00	296.0	0.0	0.0	7241.0	2566.0	174.0	0.0	0.0	0.0
35061	2018-12-31 21:00:00+01:00	292.0	0.0	0.0	7025.0	2422.0	168.0	0.0	0.0	0.0
35062	2018-12-31 22:00:00+01:00	293.0	0.0	0.0	6562.0	2293.0	163.0	0.0	0.0	0.0
35063	2018-12-31 23:00:00+01:00	290.0	0.0	0.0	6926.0	2166.0	163.0	0.0	0.0	0.0

35064 rows × 29 columns

