

Modelo inteligente para la predicción de datos a dos años vista

Nerea Solís Castrillón
Ciencia e Ingeniería de datos

Cátedra TotalEnergies de Analítica de Datos e Inteligencia Artificial



Introducción

En el sistema eléctrico español, cada vivienda está identificada mediante un código conocido como CUPS (Código Universal de Punto de Suministro), por lo que en la base de datos de la empresa la cantidad de CUPS es equivalente a la cantidad de clientes.

Otro concepto importante en el sistema es el de comercializadora, siendo esta la encargada de vender la electricidad y/o gas a los clientes finales. Su responsabilidad se basa en comprar la suficiente energía para abastecer a todos sus clientes y pagar el servicio de transporte realizado por la empresa de distribución; la factura que engloba los cobros relacionados con los servicios prestados por la distribuidora se conoce como peaje.

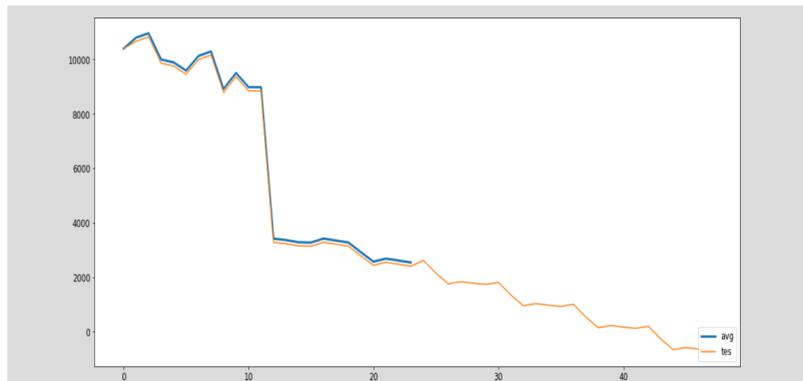
Cabe destacar la diferencia entre el importe de activa (generación de la energía), el importe de potencia (relacionado con la potencia contratada) y el importe total.

Objetivos del estudio

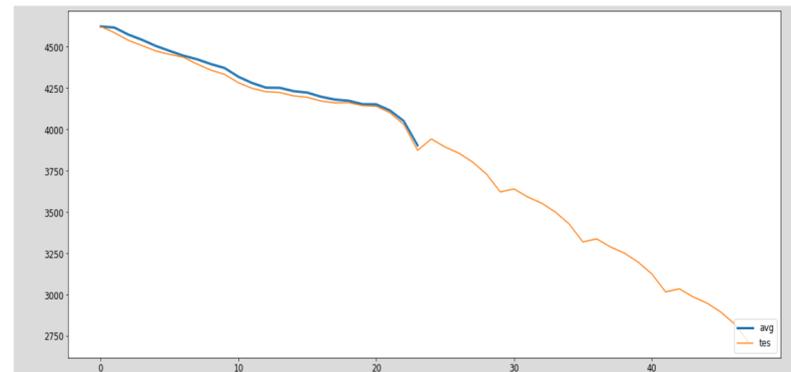
Se busca prever, con el menor error posible, la cantidad de CUPS, el consumo y el importe total, activa y de potencia para los veinticuatro meses siguientes en base a la información de los meses anteriores.

Métodos

En este proyecto se intenta desarrollar un modelo predictivo mediante el suavizado exponencial simple. En primer lugar, se exportaron los datos necesarios de la base de datos mediante la herramienta SAS y, tras usar el modelo ideado para realizar la previsión, se exportaron los datos previstos a varios archivos Excel.



Figuras y Resultados



```
Results of Dickey-Fuller Test:
Test Statistic      1.250461
p-value            0.996317
#Lags Used         9.000000
Number of Observations Used 14.000000
Critical Value (1%) -4.012034
Critical Value (5%) -3.104184
Critical Value (10%) -2.690987
dtype: float64
```

```
Results of Dickey-Fuller Test:
Test Statistic      0.000000
p-value            0.958532
#Lags Used         0.000000
Number of Observations Used 11.000000
Critical Value (1%) -4.223238
Critical Value (5%) -3.189369
Critical Value (10%) -2.729839
dtype: float64
```

Tomando como referencia la gráfica principal, en la que se representan la cantidad de CUPS a futuro para la provincia de Asturias, se puede ver que la estimación parece razonablemente acertada, sin embargo, por los resultados del test de Dickey-Fuller, se puede confirmar que el modelo no predice los datos de una manera totalmente eficaz, lo que puede deberse a la escasez de datos para poder entrenar y evaluar el modelo antes de enfrentarlo al problema real.

1	Mes	Datos
2	Agosto	329466,9
3	septiembre	327573,2
4	Octubre	325997,4
5	Noviembre	323894,9
6	Diciembre	320940,1
7	Enero	318941
8	Febrero	318753,6
9	Marzo	316859,9
10	Abril	315284,2
11	Mayo	313181,7
12	Junio	310226,8
13	Julio	308227,7
14	Agosto	308040,4
15	septiembre	306146,7
16	Octubre	304570,9
17	Noviembre	302468,4
18	Diciembre	299513,5
19	Enero	297514,4
20	Febrero	297327,1
21	Marzo	295433,4
22	Abril	293857,6
23	Mayo	291755,1
24	Junio	288800,3
25	Julio	286801,2

Como se ve en la imagen de la izquierda, los resultados obtenidos con el modelo y que sirven de previsión para los siguientes 2 años, son exportados a un archivo Excel en el que cada hoja agrupa las previsiones para los diferentes campos; CUPS, importe total, importe activa e importe de potencia.

La exportación de los datos a un Excel, que realiza de manera automática el código que contiene el modelo, facilita la presentación de los datos a la hora de realizar un análisis exhaustivo del futuro de la empresa.

Conclusiones

Se obtienen unos resultados francamente interesantes, aunque esto no implica que sean fiables al cien por cien. Como se esperaba, el modelo es capaz de adaptarse a los datos que se le dan para entrenar de cara al problema real, sin embargo, esta cantidad es insuficiente para poder entrenar el modelo y evaluarlo de manera previa a realizar la previsión que se llevará a cabo para los meses siguientes, por lo que eso da pie a que se puedan producir graves errores en varias de las previsiones.

Debido a las variaciones de los datos a lo largo de los meses anteriores, concretamente los 24 meses previos al estudio, el modelo se centra demasiado en intentar parecer su curva (curva azul en la gráfica) a la curva real para esos meses atrás (curva amarilla en la gráfica), lo que influye negativamente a la hora de enfrentarse al problema real; si el modelo adapta de manera excesiva su previsión a los datos que se dan como ejemplo para que realice el proceso de aprendizaje, al no tener datos suficientes para realizar una comprobación de prueba es muy probable que la previsión que se realice para los meses futuros cometa variedad de errores, pues durante los meses que se han cogido de referente pueden ocurrir anomalías, que como su nombre indica no se espera que sucedan de manera continua en todos los meses, por lo que, aunque el modelo tomará esas anomalías como el comportamiento normal de los datos, la previsión fallará y se llegarán a cometer errores que dependerán, en gran medida, de la cantidad de anomalías que se hayan detectado en los datos de prueba, pues a mayor cantidad de anomalías mayor será el fallo de la previsión.

Trabajo futuro

De manera general, se debe modificar el modelo usado para la previsión, pues si bien es cierto que en algunas gráficas parece que el error es mínimo, en otras hay graves errores en la previsión. Además se deben conseguir más datos para poder realizar previsiones más seguras.

Otra tarea para realizar es la creación de bucles dentro del código con el fin de que, ejecutando el código una sola vez, se realice la previsión de los cups, los importes y el consumo para todas las posibles combinaciones de los datos (todas las provincias con todas las comercializadoras junto a todas las distribuidoras).



Universidad de Oviedo