

Context

Fa uns anys vaig treballar a **Normawind**, una empresa consultora d'energia eòlica. Quan vaig començar, l'any 2013, em van assignar la tasca de desenvolupar un model de predicció de preus del mercat elèctric minorista basant-me en els preus del mercat elèctric majorista. Els preus del mercat majorista es publiquen unes hores abans, per la qual cosa una eina d'aquest tipus que pogués proporcionar amb antelació els preus del mercat minorista podia proporcionar un avantatge competitiu per alguns clients interessats.

Resolució inicial del problema

En aquell moment, per realitzar una predicció bàsica vaig decidir aplicar un ajust de regressió lineal. Comptava amb dades de tot un any, i aquest model s'ajustava raonablement bé a les dades disponibles. Tot i que la programació es realitzava en llenguatge Python, vaig optar per utilitzar equacions clàssiques per calcular els coeficients de la regressió, ja que en aquella època no existien frameworks de desenvolupament tan avançats com els actuals.

Següent pas: model de predicció per a l'endemà

El següent repte consistia a desenvolupar un model predictiu capaç d'estimar els preus del mercat elèctric minorista per a l'endemà. Tanmateix, aquest desenvolupament era molt més complex. La variació diària de preus depèn de múltiples factors externs al mercat, com ara condicions meteorològiques, variables socioeconòmiques o incidències importants en la xarxa elèctrica. Aquest conjunt de factors introdueix una forta component no lineal en els preus de l'energia, fet que exigeix algoritmes més sofisticats que una simple regressió lineal o polinòmica.

Donada la gran quantitat de dades disponibles com a possibles característiques del model, el primer pas seria preparar-les adequadament: netejar-les, estandarditzar-les i escalar-les. Un cop realitzat aquest tractament, seria raonable començar amb un algoritme senzill però capaç de capturar relacions no lineals, com el **Random Forest**. Aquest mètode és molt versàtil gràcies a la seva capacitat per identificar relacions complexes entre variables i la seva robustesa en dades heterogènies.

En cas que els resultats inicials no fossin satisfactoris, es podria avançar cap a tècniques més avançades, com ara **XGBoost** o una **xarxa neuronal artificial**.

1. **XGBoost** és un mètode basat en gradient boosting àmpliament utilitzat en competicions de predicció, conegut per la seva capacitat per manejar dades amb estructures complexes.
2. Les **xarxes neuronals artificials** són aproximadors universals, amb una gran capacitat per modelar relacions molt complexes entre característiques i etiquetes, especialment eficients quan es disposa de grans volums de dades.

Tot i això, aquests dos models tenen requisits més exigents. Tant XGBoost com les xarxes neuronals depenen fortament de l'ajust dels hiperparàmetres i requereixen una potència computacional significativa, especialment durant el procés d'entrenament.