

Proyecto 2 – Machine Learning

Implementaremos modelos de machine learning

El objetivo de este proyecto es entender la abismal diferencia que existe entre Excel y Python/R para ciencia de datos.

Esta diferencia la veremos más precisamente desde la arista de machine learning (modelos predictivos).

Comenzaremos implementando el proyecto en Excel, para entender claramente sus ventajas y limitaciones, para luego ir a Python/R y demostrar lo poderosas que son estas herramientas.

Finalmente, es importante entender que este proyecto es solo un ejemplo dentro de una infinidad de opciones (No tenemos Límites).

No olvidar que, en ciencia de datos, machine learning (modelos predictivos) es un apartado relevante, no obligatorio, pero sí relevante.

Proyecto 2 – Machine Learning

Caso por resolver

El caso que abordaremos será la implementación de un modelo de regresión, es decir, trataremos de predecir el valor de una variable numérica.

Específicamente, intentaremos estimar el salario de una persona en base a otras variables, como por ejemplo años de experiencia, años de educación, etc.

El análisis que realizaremos es extrapolable a cualquier caso real, por lo que tanto el análisis como los códigos te servirán profesionalmente.

Proyecto 2 – Machine Learning

Consideraciones

Lo normal es tomar todos los datos y separarlos en entrenamiento y testeo. Con los datos de entrenamiento entrenamos el modelo. Con los datos de testeo evaluamos si el modelo es bueno o malo.

En la práctica se suele utilizar múltiples modelos para hacer la estimación, y luego se elige el mejor en base a las métricas obtenidas (Ej: R^2)

En Excel, mediante el uso de complementos, podemos implementar un tipo de modelo llamado “regresión lineal simple/múltiple”.

El problema de Excel es que no tiene diversos modelos y no está optimizado para machine learning (lo entenderás cuando lo implementemos en R/Python)

No profundizaremos en entender en detalle el modelo de regresión, ya que existe una enorme cantidad de modelos y supuestos, pero sí nos enfocaremos en entender todo el funcionamiento de machine learning.

Vamos a Excel



Vamos Excel