

Hoja informativa: Recopilación de datos

Práctica

```
# Validación cruzada
# model: modelo no entrenado para validación cruzada;
# cv: número de bloques para validación cruzada (son 3, por defecto).

from sklearn.model_selection import cross_val_score
cross_val_score(model, features, target, cv=3)
```

Teoría

Un rastreador o raspador es un software que se utiliza para extraer datos de sitios web.

Datos sin etiquetar son datos que carecen del valor objetivo.

Etiquetado de datos o anotación de datos es el proceso de determinar los valores objetivo.

Datos etiquetados son datos con un valor objetivo conocido.

Fuga de información es una situación en la que se filtra ****información sobre el objetivo accidentalmente en las funciones.

La validación cruzada es un método de entrenamiento y prueba de modelos cuando el conjunto de entrenamiento se divide en K bloques iguales. En cada una de las etapas K, el índice de bloque i se utiliza para la validación y el resto para el entrenamiento.