



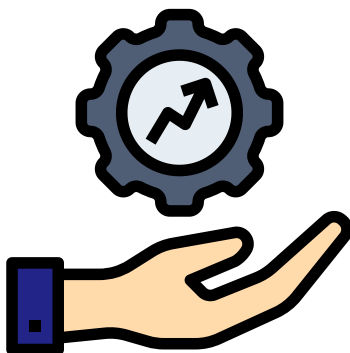
# CIENCIA DE • DATOS •

# CIENCIA DE • DATOS •

## Optimización de modelos: ajuste de hiperparámetros, selección de características.

La optimización de modelos es una etapa crítica en el proceso de construcción de modelos en Ciencia de Datos. Consiste en ajustar los hiperparámetros del modelo y seleccionar las características más relevantes para mejorar su rendimiento y capacidad predictiva. Esta etapa permite encontrar la configuración óptima del modelo y maximizar su precisión y generalización.

En esta introducción, exploraremos la importancia de la optimización de modelos, así como algunas técnicas comunes utilizadas, como el ajuste de hiperparámetros y la selección de características.



# CIENCIA DE • DATOS •

## Importancia de la optimización de modelos:

La optimización de modelos es esencial para garantizar un rendimiento óptimo y una mayor precisión en las predicciones. Los hiperparámetros son configuraciones ajustables del modelo que afectan su capacidad de aprender y generalizar a nuevos datos. Por otro lado, la selección de características implica identificar y utilizar solo las características más relevantes para el modelo, eliminando aquellas que no contribuyen significativamente a la predicción. Al optimizar los hiperparámetros y seleccionar características relevantes, se puede mejorar la eficiencia, el rendimiento y la interpretabilidad del modelo.





## Ajuste de hiperparámetros:

Los hiperparámetros son valores configurables que determinan cómo se comporta el modelo durante el entrenamiento. Ajustar los hiperparámetros adecuadamente es fundamental para obtener un rendimiento óptimo del modelo. Esto se logra a través de técnicas como la búsqueda de cuadrícula (grid search) y la búsqueda aleatoria (random search), donde se exploran diferentes combinaciones de valores de los hiperparámetros y se evalúa el rendimiento del modelo para cada combinación.

### Por ejemplo

supongamos que estamos construyendo un modelo de clasificación utilizando un algoritmo de bosques aleatorios (random forests). Los hiperparámetros que se pueden ajustar incluyen el número de árboles en el bosque, la profundidad máxima de los árboles y la cantidad mínima de muestras requeridas para dividir un nodo. Utilizando técnicas de ajuste de hiperparámetros, podemos evaluar el rendimiento del modelo con diferentes combinaciones de estos valores y seleccionar los que proporcionen el mejor rendimiento en función de las métricas de evaluación establecidas.

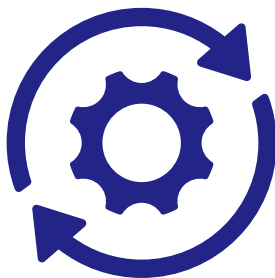
# CIENCIA DE • DATOS •

## Selección de características:

La selección de características implica identificar y utilizar solo las características más relevantes para el modelo, descartando aquellas que no aportan información significativa. Esto se logra utilizando técnicas como la correlación, la importancia de variables o algoritmos específicos, como el algoritmo LASSO (Least Absolute Shrinkage and Selection Operator) o la eliminación recursiva de características.

### Por ejemplo

si estamos construyendo un modelo de predicción de enfermedades cardíacas, podríamos tener características como la edad, el género, los niveles de colesterol, la presión arterial y los antecedentes familiares. Utilizando técnicas de selección de características, podemos identificar las características más influyentes en la predicción y eliminar las que no aporten valor. Esto puede resultar en un modelo más eficiente y preciso al considerar solo las características más relevantes.



# CIENCIA DE DATOS

La optimización de modelos es una etapa esencial en el proceso de construcción de modelos en Ciencia de Datos. Permite ajustar los hiperparámetros del modelo y seleccionar las características más relevantes para mejorar su rendimiento y capacidad predictiva. El ajuste de hiperparámetros implica explorar diferentes combinaciones de valores para encontrar la configuración óptima del modelo.

La selección de características implica identificar y utilizar solo las características más relevantes, eliminando aquellas que no aportan información significativa. Al optimizar los hiperparámetros y seleccionar características relevantes, se puede mejorar la eficiencia, el rendimiento y la interpretabilidad del modelo.

Es fundamental utilizar técnicas adecuadas, como la búsqueda de cuadrícula, la búsqueda aleatoria o algoritmos específicos, para lograr una optimización efectiva y garantizar un rendimiento óptimo del modelo en el campo de la Ciencia de Datos.

