

## 8. Selección de características (Feature Engineering)

# ¿En qué consiste la selección de características de un modelo de machine learning?

```
X = df[["variable_1", "variable_2"]]
```

```
y = df["variable_objetivo"]
```

Respuesta/=

La selección de características en un modelo de machine learning implica identificar y elegir las variables o características más relevantes para predecir la variable objetivo. En el código que proporcionaste, parece que estás seleccionando dos características, "variable\_1" y "variable\_2", del DataFrame df para usarlas como entrada (X) en tu modelo de machine learning, mientras que "variable\_objetivo" se utiliza como la variable objetivo (y).

La selección de características es importante por varias razones:

**Reducción de dimensionalidad:** Al seleccionar las características más relevantes, puedes reducir la dimensionalidad de los datos, lo que puede mejorar la eficiencia computacional y reducir el riesgo de sobreajuste.

**Mejora del rendimiento del modelo:** Al eliminar características irrelevantes o redundantes, puedes mejorar la precisión y la generalización del modelo.

**Interpretación del modelo:** Al seleccionar características significativas, el modelo resultante es más fácil de interpretar y entender.

Existen diversas técnicas para realizar la selección de características, que van desde métodos estadísticos simples hasta algoritmos más avanzados. Algunos ejemplos de técnicas comunes incluyen:

**Prueba estadística univariante:** Evalúa la relación entre cada característica y la variable objetivo utilizando pruebas estadísticas como la prueba t o la prueba F.

**Selección de características basada en modelos:** Entrena un modelo y utiliza su rendimiento para evaluar la importancia de cada característica. Algunos ejemplos son la importancia de características en árboles de decisión o coeficientes en modelos lineales.

**Selección de características recursivas:** Entrena un modelo y elimina iterativamente las características menos importantes, repitiendo el proceso hasta que se alcanza un criterio de parada.

**Selección de características basada en la información mutua:** Calcula la cantidad de información que una característica proporciona sobre la variable objetivo.

La elección de la técnica de selección de características adecuada depende del problema específico y de las características de los datos. Es importante experimentar con diferentes

métodos y evaluar su rendimiento utilizando técnicas de validación cruzada u otras métricas de evaluación del modelo.

---

---

## #11 Evaluación del modelo

# ¿Qué son las métricas de rendimiento de un modelo de Machine Learning?

```
score_train = modelo.score(X_train, y_train)
score_test = modelo.score(X_test, y_test)
print("Precisión en entrenamiento:", score_train)
print("Precisión en prueba:", score_test)
```

Respuesta/=Las métricas de rendimiento en un modelo de machine learning son medidas utilizadas para evaluar qué tan bien está funcionando el modelo en la tarea para la que fue entrenado. En el código que proporcionaste, se están utilizando dos métricas comunes para evaluar el rendimiento del modelo: precisión en entrenamiento y precisión en prueba.

Precisión en entrenamiento: Esta métrica indica qué tan bien se ajusta el modelo a los datos de entrenamiento. Se calcula comparando las predicciones del modelo con las etiquetas reales en el conjunto de datos de entrenamiento.

Precisión en prueba: Esta métrica indica qué tan bien generaliza el modelo a datos que no ha visto durante el entrenamiento, es decir, el conjunto de datos de prueba. Se calcula comparando las predicciones del modelo con las etiquetas reales en el conjunto de datos de prueba.

La precisión es solo una de las muchas métricas de rendimiento disponibles en machine learning. Otras métricas comunes incluyen la sensibilidad, la especificidad, el valor F1, el área bajo la curva ROC (AUC-ROC), entre otras, dependiendo del tipo de problema y de las características específicas de los datos.

Es importante recordar que las métricas de rendimiento deben interpretarse en el contexto del problema y las necesidades del negocio. Además, es fundamental evaluar el modelo utilizando múltiples métricas y realizar una validación cruzada para obtener una evaluación más completa del rendimiento del modelo.