

Taller



Inteligencia artificial aplicada al sector **financiero**



¿Qué **aprenderemos**?

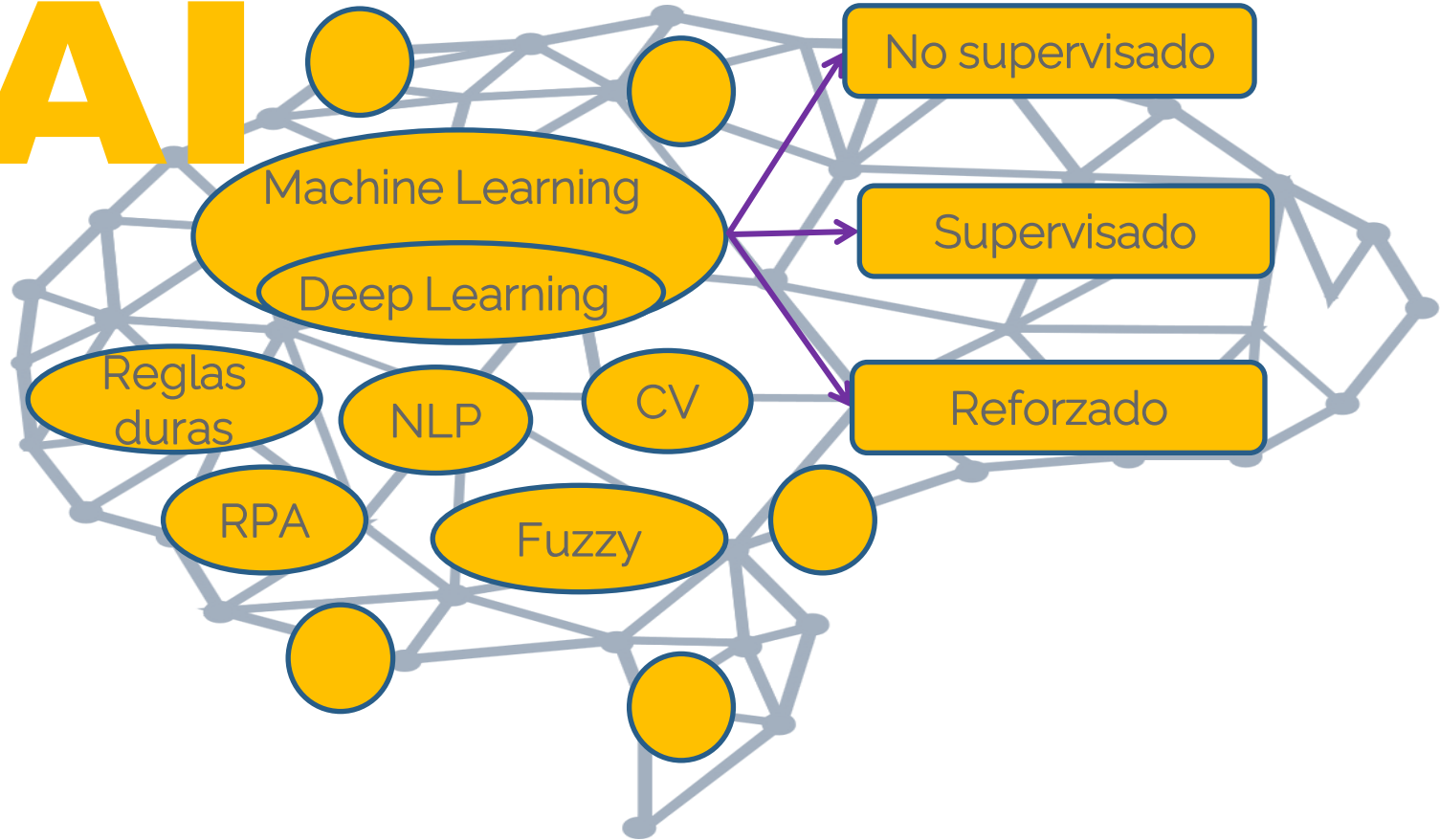
- Conceptos básicos de la Inteligencia Artificial
- Áreas clave donde la IA puede tener un impacto significativo en las operaciones financieras.
- Principales técnicas y algoritmos de IA utilizados en el sector financiero.
- Casos de uso reales de IA en finanzas.
- Habilidades básicas para implementar soluciones de IA



Inteligencia artificial en el sector financiero



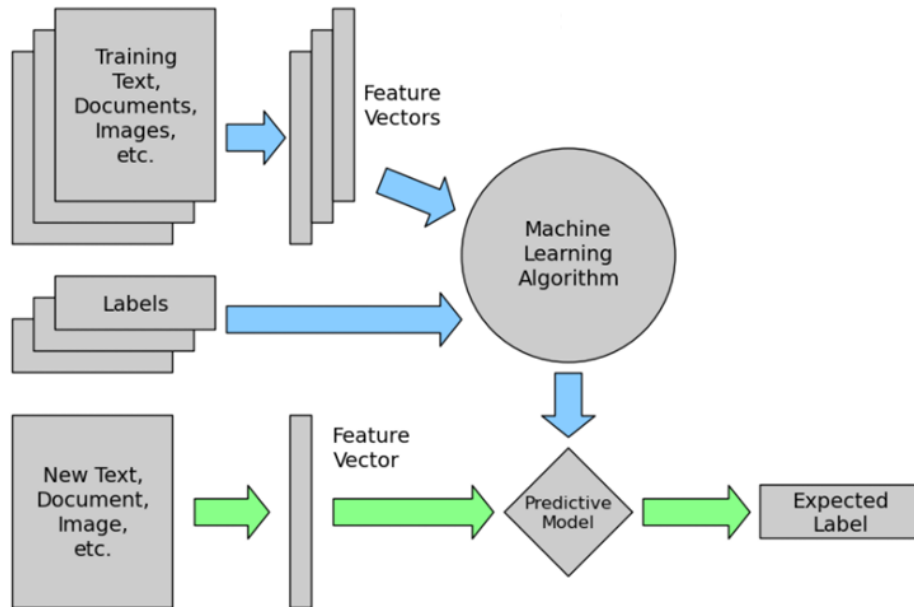
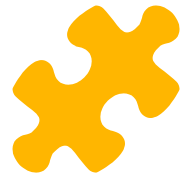
AI





Aprendizaje supervisado

Aprendizaje supervisado



Fuente: https://www.researchgate.net/publication/323107291_Context-sensitive_and_keyword_density-based_supervised_machine_learning_techniques_for_malicious_webpage_detection

Aprendizaje supervisado



Clasificación

- **Detección de fraudes:** Utilizando datos históricos de transacciones, un modelo de clasificación puede etiquetar transacciones como legítimas o fraudulentas, ayudando a las instituciones financieras a identificar patrones de fraude y tomar medidas preventivas. Regresión: crecimiento poblacional, estimación del tiempo de vida, pronósticos del mercado y pronósticos del clima.
- **Análisis de sentimientos:** En el ámbito de inversiones, un modelo puede clasificar noticias y comentarios relacionados con empresas como positivos, negativos o neutros, lo que ayuda a los inversores a evaluar el sentimiento del mercado hacia una empresa.

Aprendizaje supervisado



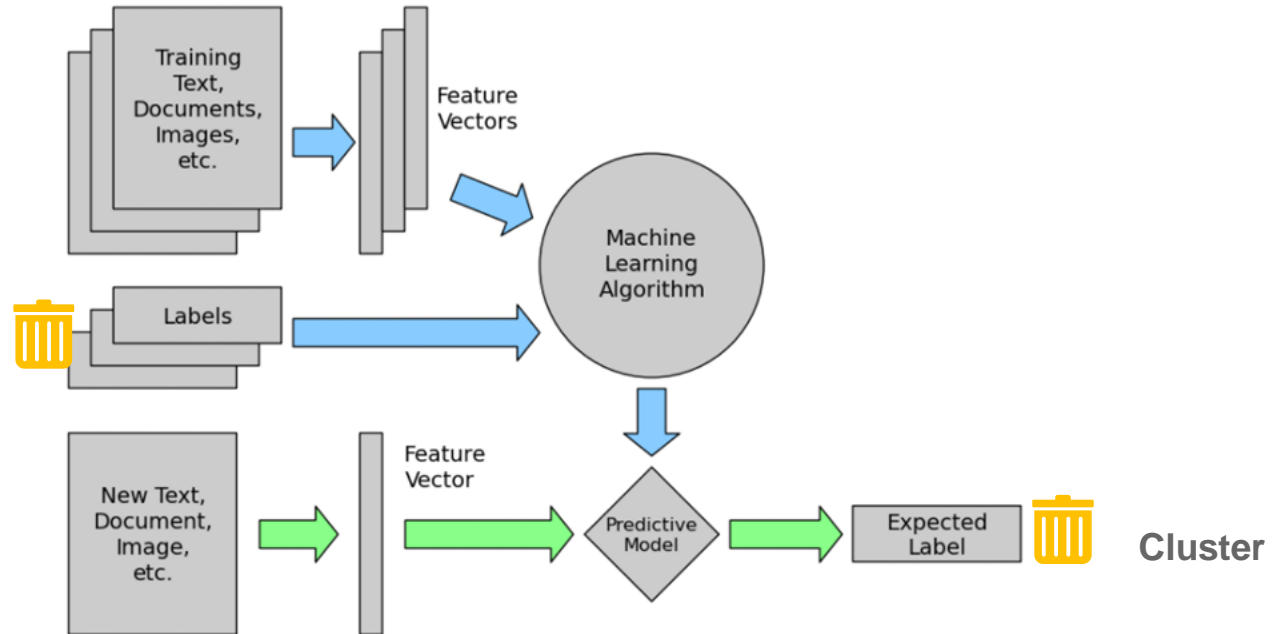
Regresión

- **Pronósticos del mercado:** Utilizando datos históricos de precios y factores económicos, un modelo de regresión puede predecir el precio futuro de una acción o un índice, lo que ayuda a los inversores a tomar decisiones informadas.
- **Modelos de riesgo crediticio:** Utilizando datos históricos de crédito y características del solicitante, un modelo de regresión puede estimar la probabilidad de impago de un préstamo, lo que ayuda a los prestamistas a evaluar el riesgo crediticio.



Aprendizaje no supervisionado

Pipeline modelo ML



Fuente: https://www.researchgate.net/publication/323107291_Context-sensitive_and_keyword_density-based_supervised_machine_learning_techniques_for_malicious_webpage_detection

Aprendizaje no supervisado



Agrupamiento (Clustering)

- **Segmentación de clientes:** Mediante técnicas de agrupamiento, las instituciones financieras pueden identificar grupos de clientes con características y comportamientos similares, permitiendo estrategias de marketing personalizadas.
- **Detección de anomalías en transacciones:** Al analizar patrones en transacciones, se pueden detectar anomalías que podrían indicar actividad fraudulenta o inusual.

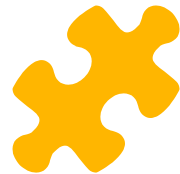
Aprendizaje no supervisado



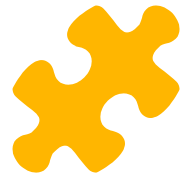
Reducción de Dimensionalidad

- **Selección de cartera:** Al reducir la dimensionalidad de los datos financieros, se pueden seleccionar las inversiones más relevantes para construir una cartera diversificada y eficiente.
- **Análisis de riesgos:** La reducción de dimensionalidad puede ayudar a identificar las variables más influyentes en la determinación del riesgo crediticio de un individuo o empresa.

Pipeline **modelo** ML



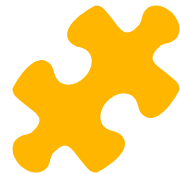
Pipeline **modelo** ML



Pipeline **modelo** ML



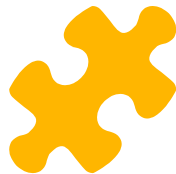
Pipeline **modelo** ML



Pipeline **modelo** ML



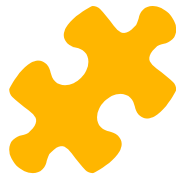
Algoritmos aprendizaje supervisado



Ejemplos de clasificación

- Regresión Logística
- Máquinas de Soporte Vectorial (SVM)
- Árboles de Decisión
- Random Forest
- Gradient Boosting (por ejemplo, XGBoost)

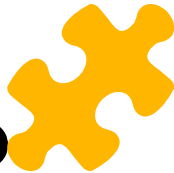
Algoritmos aprendizaje supervisado



Ejemplos de regresión

- Regresión Lineal
- Regresión Ridge y Lasso
- Máquinas de Soporte Vectorial para Regresión
- Gradient Boosting para Regresión (por ejemplo, XGBoost)
- Redes Neuronales

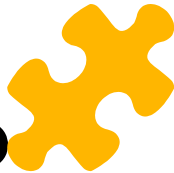
Algoritmos aprendizaje no supervisado



Ejemplos de Clustering

- K-Means
- DBSCAN
- Fuzzy C means
- Hierarchical Clustering
- Gaussian Mixture Models (GMM)

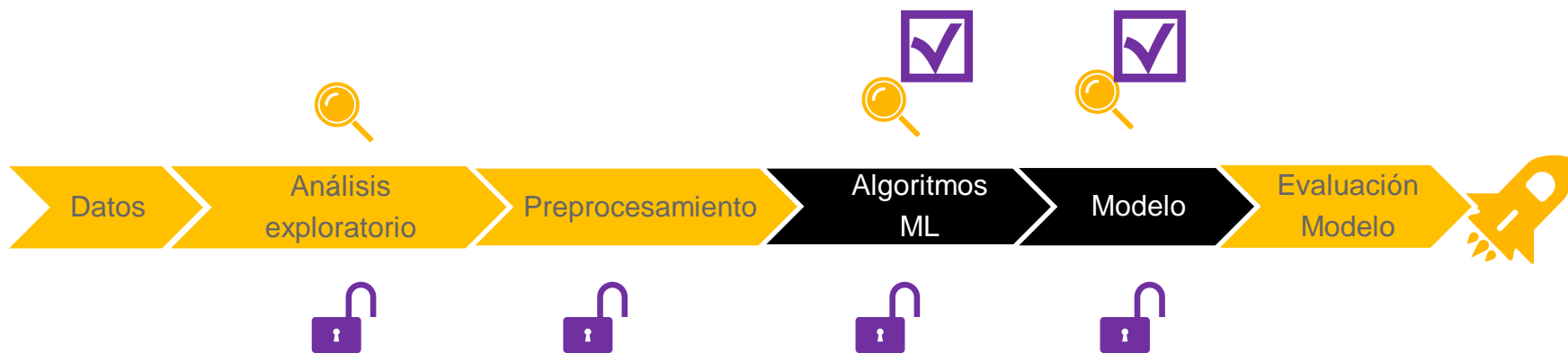
Algoritmos aprendizaje no supervisado



Ejemplos de Reducción de Dimensionalidad

- Análisis de Componentes Principales (PCA)
- Autoencoders
- t-Distributed Stochastic Neighbor Embedding (t-SNE)
- Linear Discriminant Analysis (LDA)
- Non-Negative Matrix Factorization (NMF)

Pipeline **modelo** ML



Ejemplo de técnicas de EDA

- **Visualización de Datos:** Comprender la distribución de los datos, identificar patrones, tendencias y anomalías.
- **Histogramas y Gráficos de Densidad:** Visualizar la distribución de una variable para evaluar la simetría, la presencia de valores atípicos y la forma de los datos.
- **Diagramas de Dispersión:** Explorar la relación entre dos variables, identificar correlaciones o patrones de dispersión.

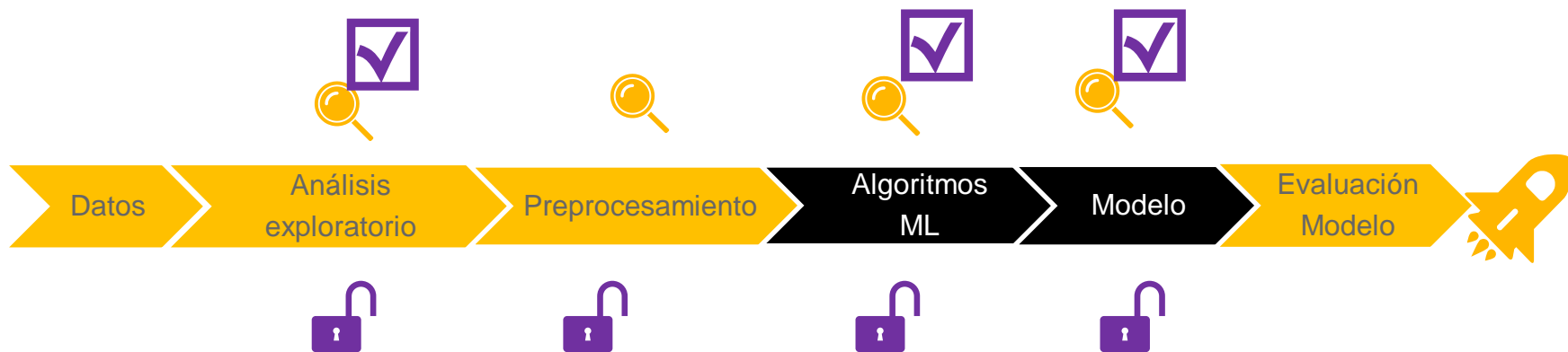
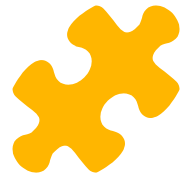
Ejemplo de técnicas de EDA

- **Diagramas de Caja (Boxplots):** Identificar valores atípicos, evaluar la dispersión y distribución de los datos.
- **Gráficos de Barras y Gráficos Circulares:** Representar frecuencias de categorías en variables categóricas.
- **Matriz de Correlación:** Evaluar las correlaciones entre pares de variables para identificar relaciones lineales.

Ejemplo de técnicas de EDA

- Diagramas de Calor (Heatmaps)
- Análisis de Componentes Principales (PCA)
- Análisis de Clústeres
- Estadísticas Descriptivas
- Gráficos de Series Temporales
- ...

Pipeline **modelo** ML



Preprocesamiento



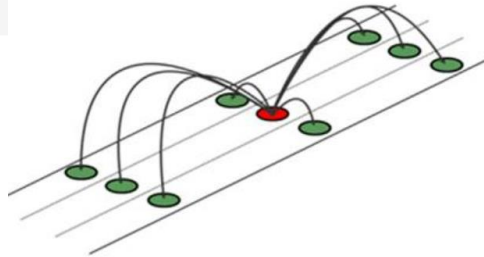
- **Imputación:** Manejar las entradas de datos faltantes para evitar problemas en el modelado y análisis posterior.

Medida	Month	Day	Ozone	Solar.R	Temp	Wind
0	5	1	41.0	190.0	67.0	7.4
1	5	2	36.0	118.0	72.0	8.0
2	5	3	12.0	149.0	74.0	12.6
3	5	4	18.0	313.0	62.0	11.5
4	5	5	NaN	NaN	56.0	

```
print(oz_mean)
```

42.12931034482759

Medida	Month	Day	Ozone	Solar.R	Temp	Wind
0	5	1	41.00000	190.0	67.0	7.4
1	5	2	36.00000	118.0	72.0	8.0
2	5	3	12.00000	149.0	74.0	12.6
3	5	4	18.00000	313.0	62.0	11.5
4	5	5	42.12931	NaN	56.0	14.3





Preprocesamiento

- Codificación de Variables Categóricas: Convertir variables categóricas en formato adecuado para su uso en algoritmos de Machine Learning.

Variable
Categoría 1
Categoría 2
Categoría 3

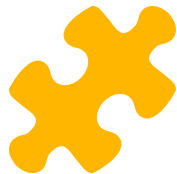


Categoría 1	Categoría 2	Categoría 3
1	0	0
0	1	0
0	0	1



Preprocesamiento

- **Normalización y Estandarización:** Escalar las variables para asegurarse de que todas tengan un rango similar, lo que puede mejorar el rendimiento de ciertos algoritmos.
- **Estandarización:**
 - Es un paso requerido por algunos algoritmos.
 - La variable obtiene una media cero y una desviación estándar igual a 1.
 - Es útil porque si una variable tiene una varianza más grande que otras puede hacer que los algoritmos no aprendan de las demás variables.
 - Sensible a outliers.



Preprocesamiento

Algunos modelos que requieren estandarización:

- KNN
- K - means
- Regresión logística
- SVM
- PCA

$$\mu = 0 \quad \sigma = 1$$

$$\mu = \frac{1}{N} \sum_{i=1}^N (x_i) \quad \sigma = \sqrt{\frac{1}{N} \sum_{i=1}^N (x_i - \mu)^2}$$

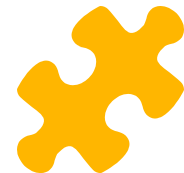
$$\overrightarrow{x_{new}} = \frac{\vec{x} - mean(\vec{x})}{std(\vec{x})}$$



Preprocesamiento

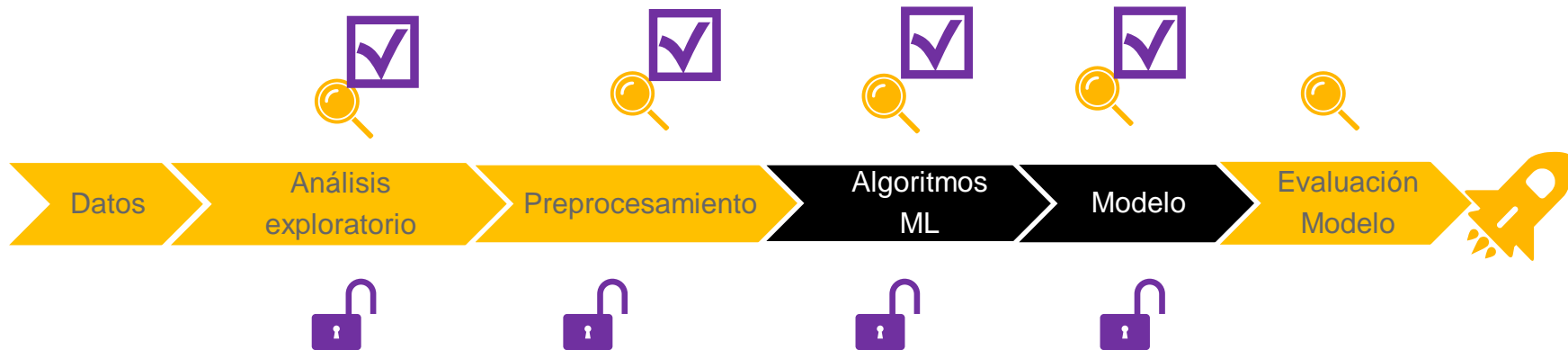
- **Eliminación de Valores Atípicos:** Identificar y tratar los valores extremos que podrían afectar negativamente el rendimiento de los modelos.
- **Transformación de Variables:** Aplicar funciones matemáticas o estadísticas para transformar las variables y ajustar su distribución.
- **Binning (Agrupación de Valores):** Agrupar valores numéricos en intervalos discretos para simplificar el análisis y el modelado.

Preprocesamiento

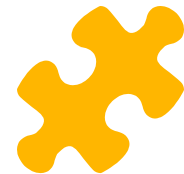


- **Reducción de Dimensionalidad:** Reducir la cantidad de variables en los datos manteniendo la información más relevante y significativa.
 - **Selección de Características:** Seleccionar las variables más importantes y relevantes para el modelado, descartando aquellas que aportan poca información.
- **Ingeniería de Características:**
 - **Extracción de Características:** crear nuevas características que sean combinaciones lineales de las características originales.
 - **Creación de Variables de Interacción:** Se pueden generar nuevas variables combinando o multiplicando características existentes para capturar relaciones no lineales.
 - ...

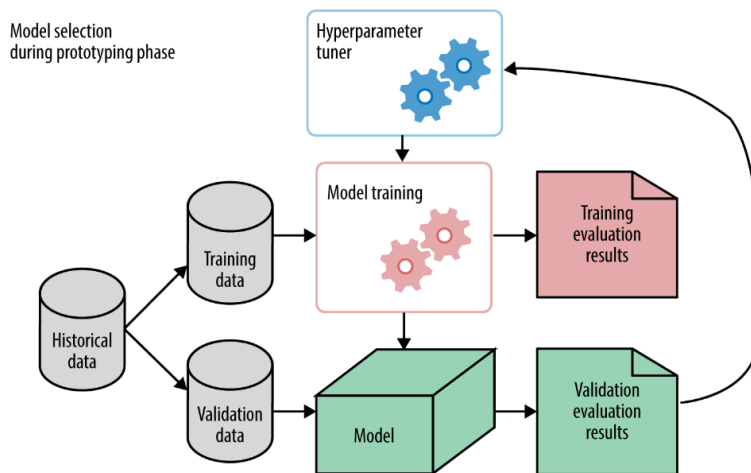
Pipeline **modelo** ML



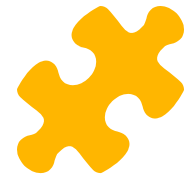
Evaluación del model



La evaluación es una etapa esencial para medir cómo un modelo se desempeña en situaciones del mundo real.



Fuente: <https://www.oreilly.com/library/view/evaluating-machine-learning/9781492048756/ch03.html>



Evaluación del model

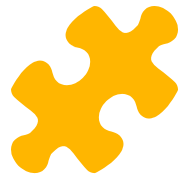
- Ajuste de Hiperparámetros
- Métricas de Desempeño
- Criterio de Generalización



Evaluación del model

Hiperparámetros

- Los hiperparámetros son configuraciones ajustables que afectan cómo el modelo se entrena y se comporta.
- Ajustar hiperparámetros de manera óptima es crucial para lograr un rendimiento óptimo.
- La validación cruzada k-fold, se puede ajustar y seleccionar hiperparámetros de manera más confiable.
- Ejemplos de hiperparámetros: tasa de aprendizaje, profundidad del árbol, número de vecinos.



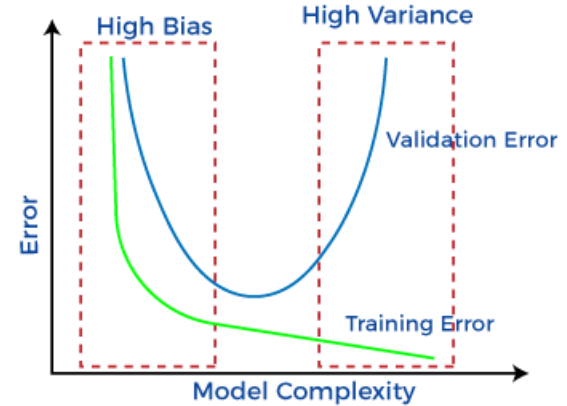
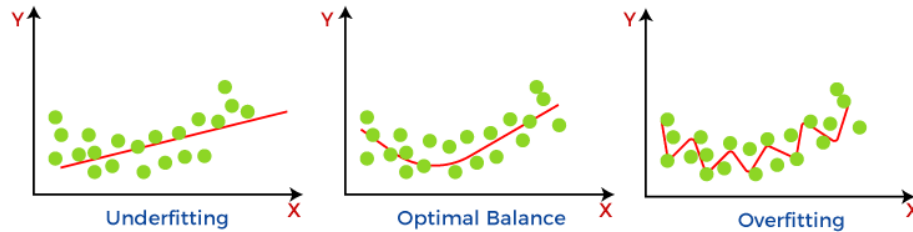
Evaluación del modelo

Métricas de desempeño

- Diferentes métricas miden diferentes aspectos del rendimiento del modelo.
- Clasificación: precisión, sensibilidad, especificidad, F1-score, área bajo la curva (AUC-ROC).
- Regresión: error cuadrático medio (MSE), raíz del error cuadrático medio (RMSE), coeficiente de determinación (R^2).

Evaluación del model

Criterio de generalización



Fuente: <https://www.analyticsvidhya.com/blog/2022/10/non-generalization-and-generalization-of-machine-learning-models/>



Vamos a los casos prácticos