

Inteligencia Artificial

Datos Etiquetados: Son datos que incluyen una etiqueta o categoría que describe lo que representan. Estas etiquetas son proporcionadas por humanos y sirven como referencia para que los modelos de aprendizaje supervisado aprendan a predecir resultados.

Ejemplo: En un conjunto de imágenes de perros y gatos, cada imagen tiene una etiqueta que indica si es un "perro" o un "gato".

Datos no etiquetados: Son datos que no tienen etiquetas o categorías asociadas. El modelo no recibe información sobre lo que representan los datos.

Ejemplo: Un conjunto de imágenes de animales sin indicar si son perros, gatos u otros.

Python & SQL relacionados con la IA

Python y la IA: Python es uno de los lenguajes de programación más populares en el campo de la IA debido a su simplicidad, flexibilidad y la gran cantidad de bibliotecas y frameworks disponibles. Aquí algunas formas en que Python se relaciona con la IA:

Bibliotecas especializadas: Python cuenta con bibliotecas como:

Tensor Flow y PyTorch para el desarrollo de modelos de aprendizaje profundo (deep learning).

Scikit-learn para algoritmos de aprendizaje automático (machine learning).

Keras para construir redes neuronales.

Pandas y NumPy para el manejo y procesamiento de datos.

Facilidad de prototipado: Python permite a los científicos de datos y desarrolladores probar ideas rápidamente, lo que es crucial en proyectos de IA.

SQL y la IA: (Structured Query Language) es el lenguaje estándar para gestionar y consultar bases de datos relacionales. Su relación con la IA se centra en el manejo y preparación de datos, que es una parte crítica en cualquier proyecto de IA. Aquí algunas formas en que SQL se relaciona con la IA:

Extracción de datos: Los modelos de IA requieren grandes cantidades de datos para entrenarse. SQL se utiliza para extraer estos datos de bases de datos relacionales.

Limpieza y preparación de datos: Antes de usar los datos en un modelo de IA, es necesario limpiarlos y transformarlos. SQL permite realizar operaciones como filtrado, agregación y unión de tablas para preparar los datos.

Almacenamiento de datos: SQL se utiliza para almacenar datos estructurados, que luego pueden ser utilizados para entrenar modelos de IA.

Integración con Python: Muchas bibliotecas de Python, como Pandas y SQLAlchemy, permiten conectarse a bases de datos SQL para leer y escribir datos directamente desde un script de Python.

Tabla de Comparación

	Aprendizaje Supervisado	Aprendizaje No Supervisado
Definición	El modelo aprende a partir de datos etiquetados (conocidos como "entradas" y "salidas").	El modelo encuentra patrones en datos no etiquetados sin referencias previas.
Objetivo	Predecir resultados o clasificar datos basándose en ejemplos proporcionados.	Descubrir estructuras ocultas o agrupaciones en los datos.
Datos requeridos	Datos etiquetados (con etiquetas o categorías conocidas).	Datos no etiquetados (sin categorías o etiquetas).
Modo de empleo	Entrenamiento con pares de entrada-salida para aprender a mapear las entradas a las salidas.	Análisis de datos para encontrar patrones, agrupaciones o reducción de dimensionalidad.
Ejemplos de modelos	<ul style="list-style-type: none">- Regresión lineal- Árboles de decisión- Redes neuronales- SVM	<ul style="list-style-type: none">- K-Means- DBSCAN- PCA- Autoencoders
Ejemplos de aplicaciones	<ul style="list-style-type: none">- Predicción de precios de viviendas- Clasificación de imágenes- Detección de spam	<ul style="list-style-type: none">- Segmentación de clientes- Detección de anomalías- Reducción de dimensionalidad
Librerías comunes	<ul style="list-style-type: none">- Scikit-learn- TensorFlow- PyTorch- Keras	<ul style="list-style-type: none">- Scikit-learn- TensorFlow- PyTorch- H2O

Ventajas	<ul style="list-style-type: none"> - Precisión alta cuando los datos están bien etiquetados. - Fácil de evaluar. 	<ul style="list-style-type: none"> - No requiere datos etiquetados. - Útil para explorar datos desconocidos.
Desventajas	<ul style="list-style-type: none"> - Requiere datos etiquetados, lo que puede ser costoso y laborioso. - Menos flexible. 	<ul style="list-style-type: none"> - Menos preciso en tareas específicas. - Difícil de evaluar resultados.

Referencias

Goodfellow, I., Bengio, Y., & Courville, A. (2016). Deep learning. MIT Press.

Mitchell, T. M. (1997). Machine learning. McGraw-Hill.

Zhu, X., & Goldberg, A. B. (2009). Introduction to semi-supervised learning. Synthesis Lectures on Artificial Intelligence and Machine Learning, 3(1), 1-130.

<https://doi.org/10.2200/S00196ED1V01Y200906AIM006>

McKinney, W. (2017). Python for data analysis: Data wrangling with pandas, NumPy, and IPython (2nd ed.). O'Reilly Media.

Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, E. (2011). Scikit-learn: Machine learning in Python. Journal of Machine Learning Research, 12, 2825–2830.

Géron, A. (2019). Hands-on machine learning with Scikit-Learn, Keras, and TensorFlow: Concepts, tools, and techniques to build intelligent systems (2nd ed.). O'Reilly Media.