

ALGORITMOS DE APRENDIZAJE AUTOMÁTICO PARA ANÁLISIS Y PREDICCIÓN DE DATOS

Katty Fuertes

08 de octubre de 2023

Resumen

Actualmente, estamos inmersos en una era de información en la que las empresas tienen la capacidad de anticipar nuestras preferencias y predecir qué productos nos gustarán en función de nuestro comportamiento en línea. Utilizan nuestros datos para crear estrategias de marketing más efectivas y reducir los riesgos al lanzar nuevos productos. Este conocimiento se obtiene a través de la Ciencia de Datos y la técnica de Aprendizaje Automático (Machine Learning). Trabajan con enormes cantidades de datos que, con el tiempo, revelan patrones que probablemente se repetirán en el futuro, lo que permite realizar predicciones precisas.

1. Introducción

En un mundo donde la cantidad de información generada en Internet es abrumadora, desde redes sociales hasta transacciones comerciales y datos de dispositivos, existe una forma de aprovechar todo este flujo de información. En lugar de simplemente almacenar estos datos en servidores, se utilizan para analizar comportamientos y prever tendencias futuras. En ocasiones, se recopila tanta información que es posible anticipar con seguridad cómo se comportará un grupo de personas o dispositivos en el futuro. Este proceso se basa en el Machine Learning, una técnica fundamental de la Ciencia de Datos, que utiliza métodos para hacer predicciones basadas en datos y patrones. En esta introducción, exploraremos el Machine Learning y cómo se aplica en el mundo de la Ciencia de Datos.

2. Inteligencia Artificial

La inteligencia artificial es una tecnología que permite a los dispositivos realizar tareas similares a las acciones humanas, como dar respuestas automáticas o resolver problemas a través de lógica formal. Estos dispositivos pueden aprender y discernir, lo que los hace casi superhumanos en términos de velocidad de

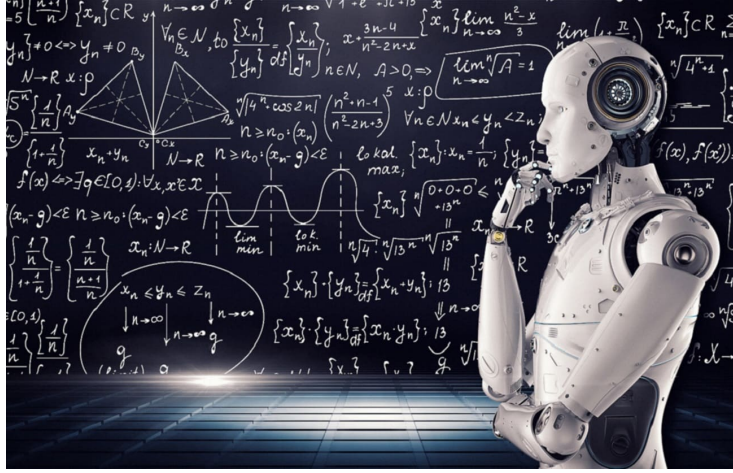


Figura 1: Inteligencia Artificial

procesamiento y funcionamiento continuo sin necesidad de descanso. Esto les otorga ventajas significativas sobre los seres vivos.

La capacidad de la inteligencia artificial para aprender y adaptarse es una de sus características más destacadas. Los dispositivos pueden mejorar su rendimiento con el tiempo, lo que los convierte en herramientas poderosas en una variedad de aplicaciones, desde asistentes virtuales hasta vehículos autónomos. La inteligencia artificial está en constante evolución y promete un futuro emocionante en el mundo de la tecnología y la automatización.

Sin embargo, a medida que la inteligencia artificial se vuelve más avanzada, también surgen desafíos éticos y de seguridad que deben abordarse. El potencial de la inteligencia artificial es innegable, pero su uso responsable y cuidadoso es esencial para garantizar un impacto positivo en la sociedad. [4]

3. Machine Learning

El aprendizaje automático, o "machine learning" en inglés, es un campo fundamental de la inteligencia artificial que se centra en el desarrollo de algoritmos y modelos que permiten a las computadoras aprender y mejorar su rendimiento en tareas específicas sin ser programadas explícitamente. En lugar de seguir instrucciones rígidas, las máquinas utilizan datos y experiencias pasadas para tomar decisiones, identificar patrones y predecir resultados futuros.

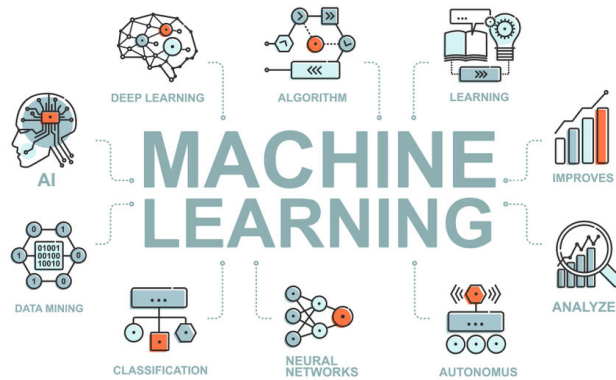


Figura 2: Machine Learning

El aprendizaje automático ha revolucionado muchas industrias al permitir la automatización de tareas, el análisis de grandes cantidades de datos y la toma de decisiones más precisas. A medida que avanza la investigación en este campo, se esperan avances continuos y una mayor integración en la vida cotidiana, desde la atención médica hasta la conducción autónoma y más. Hay dos tipos de aprendizaje: el supervisado y el no supervisado.

Aprendizaje Supervisado

El aprendizaje supervisado es un enfoque fundamental en el campo del aprendizaje automático (machine learning) que se utiliza para entrenar modelos de inteligencia artificial. En el aprendizaje supervisado, se proporciona al algoritmo de aprendizaje un conjunto de datos de entrenamiento que consta de pares de entrada y salida, donde la entrada representa características o datos de entrada y la salida es la etiqueta o resultado deseado.

El aprendizaje supervisado se basa en el uso de variables dependientes (etiquetas) y variables independientes (características) para construir modelos que puedan hacer predicciones precisas. Estas variables son esenciales en la formulación y resolución de problemas de machine learning.

- Los algoritmos de clasificación son como detectives de datos. Analizan patrones en la información que les proporcionamos y la dividen en diferentes grupos. Después, cuando les presentamos nuevos datos, los comparan con los patrones que han aprendido y los colocan en el grupo apropiado. Esto nos permite predecir a qué categoría pertenecen los nuevos elementos de

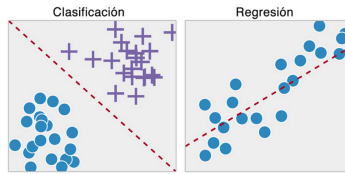


Figura 3: Clasificación y Regresión

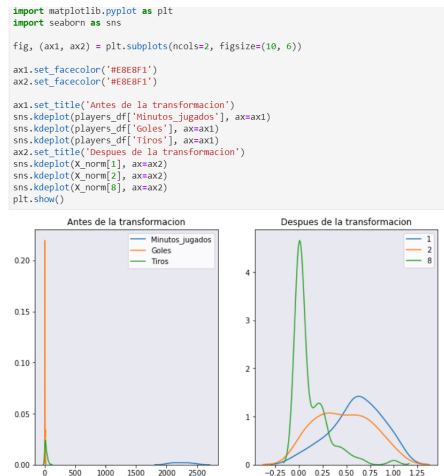


Figura 4: Aprendizaje No Supervisado

manera automática. En esencia, estos algoritmos son como asistentes que nos ayudan a organizar y etiquetar datos de manera inteligente.

- Los algoritmos de regresión En este enfoque, lo que se busca es obtener un valor numérico en lugar de clasificar en grupos. En lugar de asignar un elemento a una categoría, el método proporciona un número concreto como resultado. Es como obtener una respuesta numérica precisa en lugar de una etiqueta de grupo.

Aprendizaje No Supervisado

El algoritmo se encarga de identificar y agrupar datos en función de sus similitudes en términos de características compartidas, sin hacer referencia a etiquetas específicas. El algoritmo realiza esta tarea basándose en la presunción de que los datos que comparten ciertas características podrían pertenecer al mismo grupo.

4. Modelos de Machine Learning

Modelos Lineales

Trata de identificar una línea que se ajuste adecuadamente a un conjunto de puntos en un gráfico. Entre las opciones más comunes se encuentran la regresión lineal (también llamada regresión de mínimos cuadrados) y la regresión logística (que se adapta a problemas de clasificación cuando se manejan datos discretos o categóricos). Sin embargo, estos dos modelos enfrentan el problema del "sobreajuste", lo que significa que se ajustan en exceso a los datos disponibles, lo que puede ser perjudicial para nuevos datos que puedan surgir. Debido a su simplicidad, estos modelos no brindan resultados óptimos para situaciones más complejas.

Modelos de Árbol

Estos modelos se caracterizan por su precisión y estabilidad, además de ser más fáciles de entender porque generan reglas de decisión que se pueden visualizar como un árbol. A diferencia de los modelos lineales, tienen la capacidad de capturar relaciones no lineales para resolver problemas. Los árboles de decisión y los random forests (que son una combinación de múltiples árboles de decisión) son ejemplos destacados de estos modelos. A pesar de su mayor precisión y complejidad, es importante tener en cuenta que su rendimiento puede ser inferior en comparación con modelos más simples.

Rede Neurales

Las redes neuronales artificiales buscan imitar de cierta manera el funcionamiento del cerebro, que consta de millones de neuronas interconectadas que se comunican entre sí. Esta imitación del cerebro humano es popular debido a sus capacidades cognitivas avanzadas. Por ejemplo, son ideales para tareas como el reconocimiento de imágenes o videos, que son procesos complejos. Sin embargo, al igual que el cerebro humano, estas redes neuronales artificiales requieren un entrenamiento prolongado y una gran capacidad de procesamiento, lo que puede hacer que sean lentas en su aprendizaje. [2]

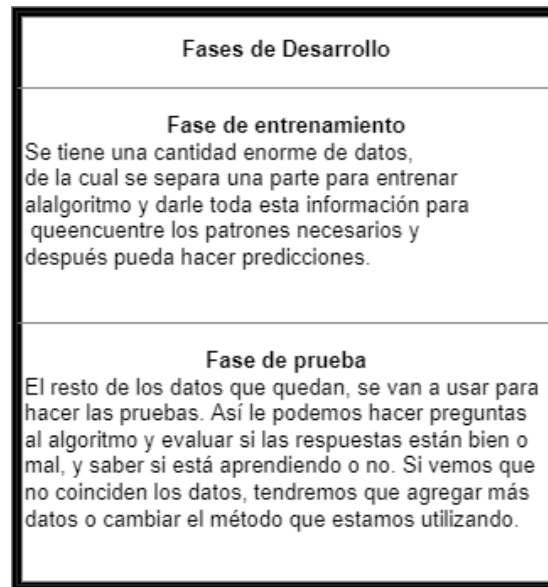


Figura 5: Fases de Machine Learning
[1]

5. Diferencias entre el Machine Learning y el Business Intelligence

Business Intelligence: Comienza recopilando datos brutos de bases de datos transaccionales, luego los transforma y carga en un DataWarehouse estructurado utilizando herramientas ETL.

Machine Learning: Se enfoca en detectar patrones a partir de datos individuales con características distintivas y utiliza miles de variables para detectar estos patrones. No se basa en la agregación de datos.

Business Intelligence: Su objetivo principal es proporcionar información histórica y actual, generalmente a través de paneles visuales y dashboards, para ayudar a entender los resultados pasados y ajustar la estrategia futura.

Machine Learning: Se centra en la creación de aplicaciones predictivas que automatizan procesos, toman decisiones y aprenden continuamente a medida que se les suministran nuevos datos. Ofrece análisis predictivos en lugar de descriptivos.

6. Conclusiones

Los algoritmos de aprendizaje automático sean efectivos en el análisis y predicción de datos, es fundamental llevar a cabo un proceso de entrenamiento adecuado y ajustar los hiperparámetros. Los modelos de aprendizaje automático deben recibir datos etiquetados para aprender patrones y hacer predicciones.

El campo de los algoritmos de aprendizaje automático está en constante evolución y los algoritmos deben ser adaptables para seguir el ritmo de los cambios en los datos y los requisitos de las aplicaciones. La adaptabilidad es fundamental para garantizar que los algoritmos sigan siendo relevantes y eficaces a medida que cambian las circunstancias y los patrones de los datos. [3]

Referencias

- [1] Conchas. 8 aplicaciones de machine learning. *8 aplicaciones de Machine Learning*, 2017.
- [2] Corinna Cortes and Vladimir Vapnik. Guía para comenzar con algoritmos demachine learning. *algoritmos Machine Learning*, 2017.
- [3] Jordi Navarro y Andrés González José Antonio Ballarín, Pere Vandellós. Diferencias entre business intelligence y machine learning. *Diferencias entre Business Intelligence y Machine Learning*, 2017.
- [4] Ana Pérez Porto, Julián y Gardey. Inteligencia artificial - qué es, definición y concepto. *Inteligencia artificial*, (1), 2011.