

Fase 1 - Evaluación Inicial

Machine Learning

O. Andrés Russi Porras.

oarussir@unadvirtual.edu.co

Universidad Nacional Abierta y a Distancia

Resumen

El documento examina el aprendizaje automático (ML) y su importancia en varios ámbitos. Define el ML como una técnica que permite a las computadoras aprender sin programación específica, resaltando su adaptabilidad y mejora con grandes conjuntos de datos. Presenta una distinción entre aprendizaje supervisado y no supervisado. Luego, analiza la relevancia del ML en la capacidad para procesar grandes volúmenes de datos, automatizar decisiones, personalizar experiencias, impulsar la innovación y desarrollar sistemas inteligentes. Concluye con una aplicación específica del ML en la automatización robótica de procesos (RPA), enfocándose en el "Document Understanding" para mejorar la extracción de datos de documentos.

Palabras clave: Aprendizaje automático, Supervisado, No supervisado, Procesamiento de datos, Automatización.

1. ¿Qué es el aprendizaje automático?

El aprendizaje automático (ML) es un subcampo de la ciencia de datos que capacita a las computadoras para adquirir habilidades y conocimientos mediante la experiencia, sin una programación específica para cada tarea individual. Esta capacidad de aprendizaje se consigue con algoritmos especializados de aprendizaje automático, diseñados para analizar y aprender de grandes conjuntos de datos.

Lo que distingue al aprendizaje automático de la programación tradicional es su enfoque en el desarrollo de sistemas que pueden adaptarse y mejorar con el tiempo. En lugar de seguir instrucciones rígidas y predefinidas, los sistemas de aprendizaje automático utilizan datos para identificar patrones y reglas, y luego aplican estos aprendizajes para tomar decisiones o realizar predicciones.

Esta capacidad de aprender de los datos es lo que hace al aprendizaje automático tan valioso en el mundo moderno, donde la cantidad de datos generados y recopilados es vasta y crece continuamente. Al ser capaces de extraer insights y conocimientos de estos datos, los sistemas de aprendizaje automático pueden realizar tareas complejas, a menudo superando la eficiencia y precisión de los seres humanos en ciertos dominios. [Su, pag 1-7]

2. ¿Qué son los aprendizaje supervisados y no supervisados?

- **Aprendizaje supervisado:** En el aprendizaje supervisado, una modalidad del aprendizaje automático, el modelo es entrenado utilizando un conjunto de datos previamente etiquetados. Aquí, el algoritmo recibe ejemplos que incluyen tanto las entradas como las correspondientes salidas deseadas. A través de este proceso, el modelo se capacita para producir la salida adecuada basándose en nuevas entradas. Un ejemplo de aprendizaje supervisado es la clasificación. En la clasificación, el modelo se entrena para asignar etiquetas a las entradas. Por ejemplo, un modelo de clasificación se puede entrenar para identificar imágenes de gatos y perros. Otro ejemplo de aprendizaje supervisado es la regresión. En la regresión, el modelo se entrena para predecir un valor continuo a partir de las entradas. Por ejemplo, un modelo de regresión se puede entrenar para predecir el precio de una casa a partir de su tamaño y ubicación.
- **Aprendizaje no supervisado:** El aprendizaje no supervisado es un tipo de aprendizaje automático en el que el modelo trabaja con datos que no están etiquetados. Esto significa que el modelo no sabe qué son las salidas deseadas. El objetivo del aprendizaje no supervisado es explorar la estructura de los datos para identificar patrones. Un ejemplo de aprendizaje no supervisado es la agrupación. En la agrupación,

el modelo se utiliza para agrupar los datos en grupos similares. Por ejemplo, un modelo de agrupación se puede utilizar para agrupar a los clientes en función de sus hábitos de compra. Otro ejemplo de aprendizaje no supervisado es la reducción de dimensionalidad. En la reducción de dimensionalidad, el modelo se utiliza para reducir el número de dimensiones de un conjunto de datos. Esto puede ser útil para simplificar el análisis de los datos. [Su, pag 19-23]

3. ¿Por qué es importante el aprendizaje automático?

El aprendizaje automático ofrece aplicaciones en una amplia gama de campos, desde la medicina hasta las finanzas, pasando por la manufactura y la investigación científica.

Las principales razones de la importancia del ML son las siguientes:

- **Capacidad de procesar grandes volúmenes de datos:** En la era de la información, se generan enormes cantidades de datos. El ML es esencial para analizar y extraer significado de estos datos, lo que sería imposible o muy costoso con métodos tradicionales.
- **Automatización de procesos de toma de decisiones:** El ML permite a las máquinas tomar decisiones basadas en datos sin intervención humana. Esto resulta en procesos más rápidos y eficientes, y reduce la posibilidad de errores humanos.
- **Personalización y mejora de la experiencia del usuario:** El ML se utiliza para personalizar las experiencias de los usuarios, como en recomendaciones personalizadas en plataformas de streaming o compras en línea. Esto mejora la satisfacción y retención del cliente.
- **Innovación en diversos sectores:** El ML tiene aplicaciones en una amplia gama de campos, impulsando la innovación y eficiencia en estos sectores. Por ejemplo, se utiliza en la medicina para diagnósticos y tratamientos personalizados, en las finanzas para la detección de fraudes, y en la manufactura para la optimización de procesos.

- **Desarrollo de sistemas inteligentes y autónomos:** El ML facilita el desarrollo de sistemas inteligentes como vehículos autónomos y robots. Estos sistemas pueden realizar tareas complejas de manera eficiente y segura.
- **Predicción y análisis de tendencias:** El ML permite predecir tendencias y comportamientos futuros. Esto es crucial para la planificación y toma de decisiones estratégicas en negocios y gobierno.
- **Contribución a la investigación científica:** El ML ayuda a los investigadores a descubrir patrones y correlaciones en grandes conjuntos de datos. Esto puede conducir a nuevos descubrimientos científicos.

En resumen, el ML es una herramienta indispensable para el mundo moderno. Ofrece herramientas poderosas para entender y utilizar grandes cantidades de datos, mejora la eficiencia y efectividad en una variedad de aplicaciones, y es un motor clave para la innovación y el avance tecnológico. [Shi, pag 1-6]

4. ¿Qué aplicaciones tiene el aprendizaje automático?

Como desarrollador de RPA, he visto cómo el aprendizaje automático está revolucionando nuestra disciplina. Una de las aplicaciones más destacadas es el "Document Understanding", o comprensión de documentos. El "Document Understanding" se refiere al uso de algoritmos de aprendizaje automático para interpretar y procesar documentos. En el contexto de RPA, esto implica extraer información relevante de documentos de distintos formatos y estructuras, como facturas, formularios, contratos, correos electrónicos, entre otros.

En mi trabajo diario, utilizo el aprendizaje automático para automatizar la extracción y el procesamiento de datos de documentos. Esto se realiza mediante modelos de aprendizaje automático entrenados para reconocer y entender el texto, las imágenes y la estructura general de los documentos. Por ejemplo, al trabajar con facturas, el modelo puede ser entrenado para identificar y extraer datos clave como el nombre del proveedor, las fechas, los montos y otros detalles relevantes.

Esta aplicación es especialmente útil porque los documentos suelen venir en formatos variados y con diferencias en su presentación. Esto hace difícil

para un sistema basado en reglas rígidas manejarlos de manera efectiva. El aprendizaje automático, en cambio, permite que el sistema aprenda de una variedad de ejemplos y se adapte a nuevos formatos o variaciones en los documentos.

Además, la integración de Document Understanding en RPA también mejora la eficiencia y precisión en el procesamiento de datos. Reduce la necesidad de intervención manual, minimizando errores y acelerando los procesos. Esto es crucial en situaciones donde el volumen de documentos es grande y la rapidez en el procesamiento es esencial.

En resumen, el aprendizaje automático aplicado en Document Understanding es una herramienta fundamental para los desarrolladores de RPA. No solo mejora la capacidad de nuestros sistemas para manejar tareas complejas y variadas de procesamiento de documentos, sino que también aumenta significativamente la eficiencia, precisión y escalabilidad de nuestras soluciones de automatización. [MyMu, pag 5-20]

Referencias

- [MyMu] Myers, P., Kavas, I., Muller, M., & Solove, J. (2015). Intelligent Document Capture with Ephesoft. Packt Publishing Ltd.
- [Shi] Shinde, P. P., & Shah, S. (2018, August). A review of machine learning and deep learning applications. In 2018 Fourth international conference on computing communication control and automation (ICCUBE). IEEE.
- [Su] Subasi, A. (2020). Practical machine learning for data analysis using python. Academic Press.