

INTELIGENCIA ARTIFICIAL Y CIENCIA DE DATOS

SANTIAGO GONZÁLEZ MEDELLÍN
GERENCIA DE INTELIGENCIA ARTIFICIAL

En la actualidad, gracias a los altos niveles de conectividad, que nos ofrece internet como, plataforma de comunicaciones; cada día se generan volúmenes de datos demasiado grandes, estos datos son generados a través de redes sociales, plataformas de compra-venta, aplicaciones móviles, etc. Para poder obtener información útil de estos datos, se han aplicado técnicas y modelos matemáticos que permiten analizar información, que es difícil de ver a simple vista, estas técnicas son parte de un área de la Inteligencia Artificial llamada ciencia de datos.

La ciencia de datos es un área multidisciplinaria que se apoya principalmente de las siguientes áreas.

- *Matemáticas*: Son la herramienta fundamental para las demás áreas, gracias a las definiciones y propiedades establecidas por esta ciencia es posible crear modelos y establecer relaciones entre los datos.
- *Estadística*: Nos permite visualizar los datos para darnos una idea de cómo se están comportando y con base en eso poder tomar decisiones.
- *Programación*: Sería muy complicado realizar el análisis de los datos de manera manual, para ello, se implementan rutinas automatizadas y programas con los que se manipulan los datos. Los lenguajes de programación son las herramientas con las que implementan todas las técnicas de ciencia de datos.
- *Aprendizaje Automatizado*: Es el área encargada de crear modelos que capturen las relaciones entre las variables de los datos. Principalmente se utiliza para tareas de clasificación y regresión.
- *Minería de datos*: Es un conjunto de técnicas que permite extraer información útil de un conjunto de datos, apoyada de las otras áreas para lograr este objetivo.

Para realizar un proceso completo de análisis de datos, un científico de datos tiene que aplicar los siguientes pasos.

1. Extracción de datos: De las diversas fuentes en las que estén distribuidas.
2. Limpieza de datos: Pre procesar los datos para eliminar aquellos que puedan provocar sesgos en el análisis y dejarlos en un formato adecuado para la siguiente fase.



3. Desarrollar modelos que permitan extraer información de estos datos o representar las relaciones entre ellos.
4. Crear visualizaciones o aplicaciones en las que se reflejen los resultados de la experimentación.
5. Evaluación continua: Monitorear el desempeño y ajustarlo periódicamente para que esté actualizado con los datos nuevos que se van generando.

Las técnicas que son tendencia, hoy en día, son las de aprendizaje automatizado (también conocido como *machine learning*). El aprendizaje automatizado se divide en 3 tipos: (aprendizaje supervisado, no supervisado y por refuerzo).

El aprendizaje supervisado, utiliza ejemplos para aprender relaciones entre los datos, a partir de estos ejemplos, puede detectar patrones y reproducir resultados similares a los de entrenamiento con datos nuevos. Por ejemplo, puede entrenarse un modelo con datos del clima, para predecir cómo va a estar el clima a partir de datos históricos.

El aprendizaje no supervisado, agrupa datos a partir de criterios de semejanza implícitamente programados en estas técnicas, se usa principalmente para segmentar poblaciones, ya sea un grupo de clientes, para poder hacer un análisis más a profundidad de ellos.

Por último, **el aprendizaje por refuerzo**, se usa para entrenar modelos que se adapten a una situación o contexto y puedan actuar de manera adecuada. Por ejemplo, se puede entrenar un robot para aprender a caminar haciendo muchas simulaciones de ello y dando una buena puntuación a los mejores movimientos.

Las grandes cantidades de datos que las empresas tienen han hecho necesario el uso de estas técnicas para beneficiarse de los resultados del análisis de datos. Algunas de estas han llevado a aplicaciones de uso común como el reconocimiento facial de Facebook, el sistema de recomendación de varios servicios de streaming como: *Netflix, Amazon, Spotify, etc.*

Por parte de los autos de Tesla Motors, analizan los datos de simulaciones de manejo para crear modelos que apoyen al piloto automático a manejar.

Del lado financiero, se tienen modelos que apoyan al análisis de riesgos, con técnicas de *machine learning*, se ha logrado identificar movimientos fraudulentos en la actividad de la tarjeta de crédito de un cliente, también se han utilizados para evaluar el riesgo de otorgarle beneficios a personas como seguros o créditos.





En la medicina, se han realizado aplicaciones que apoyan al diagnóstico de enfermedades, llegando a tener un desempeño igual o mejor a un experto humano.

Dentro de PRAXIS se utilizan este tipo de técnicas para desarrollar soluciones en áreas en las que se necesite automatizar procesos de manera inteligente, para ello se propone el binomio inteligente, en el que el área de Inteligencia Artificial, funge como apoyo para todas las demás áreas para evolucionar las aplicaciones existentes, con el uso de herramientas de ciencias de datos; se llama binomio inteligente ya que combina las habilidades técnicas de un elemento del área de Inteligencia Artificial, con alguno de otra área. Actualmente se están realizando varias propuestas para proyectos que involucren análisis de datos, por ejemplo, junto con el área de productos financieros, para un cliente del sector de retail, se propuso mejorar su cadena de distribución, optimizando la cantidad de productos con los que se va a abastecer cada punto de venta, esto se puede realizar entrenando un modelo predictivo que calcule la cantidad que se espera vender, en el siguiente periodo de algún producto dado. Por otra parte, el robot Neek se está programando para apoyar con capacitación y difusión de las actividades dentro del área con el propósito de dar a conocer al público que actualmente se tiene un área de Inteligencia Artificial.

Actualmente se están definiendo más servicios que puede aportar el área de Inteligencia Artificial, todos ellos, pensados para liberar la carga de trabajo trivial a las personas, e incrementar el valor de sus actividades, la Inteligencia Artificial con ayuda de la ciencia de datos, es la herramienta que va a impulsar a PRAXIS al siguiente nivel en cuanto a desarrollo de soluciones.

