**Fundamentos de la Inteligencia Artificial para Data y Machine Learning.**

¿Qué es IA?

La inteligencia artificial (IA) es la capacidad de las computadoras para reemplazar tareas humanas con un poder y alcance impresionantes. Esta combina ingeniería y matemáticas para asegurar una labor más eficaz mediante sistemas informáticos sólidos.

La disciplina está ampliamente presente y despierta interés en diversas industrias, teniendo un impacto en la sociedad. Además, engloba modelos, empresas, técnicas, investigación, blogs y cursos, con una capacidad de evolución y relevancia profesional.

La inteligencia artificial (IA) es la capacidad de las computadoras para reemplazar tareas humanas con un poder y alcance impresionantes. Esta combina ingeniería y matemáticas para asegurar una labor más eficaz mediante sistemas informáticos sólidos.

La disciplina está ampliamente presente y despierta interés en diversas industrias, teniendo un impacto en la sociedad. Además, engloba modelos, empresas, técnicas, investigación, blogs y cursos, con una capacidad de evolución y relevancia profesional.

**Aplicaciones y campos de la inteligencia artificial**

La Inteligencia Artificial ofrece un vasto mundo de posibilidades en diversos campos y disciplinas. Algunos de los conceptos clave que exploraremos en este curso incluyen:

* Procesamiento del lenguaje natural (NLP): la capacidad de las máquinas para entender y generar lenguaje humano.
* Visión por computadora: el análisis e interpretación de imágenes y videos por parte de las máquinas.
* Aprendizaje profundo (Deep Learning): el uso de redes neuronales para aprender y extraer información de grandes conjuntos de datos.
* Aprendizaje por refuerzo: la habilidad de las máquinas para aprender a través de la interacción con su entorno y la retroalimentación recibida.
* Estos son solo algunos de los muchos conceptos y sectores que se benefician de la Inteligencia Artificial. En este curso, profundizaremos en cada uno de ellos, proporcionándote las bases teóricas y prácticas necesarias para comprender y aplicar estos conocimientos en el mundo real.

**¿Qué es Machine Learning?**

El \*\*machine learning\*\* (ML) es una rama de la inteligencia artificial (IA) y la informática que se centra en el uso de datos y algoritmos para que la IA imite el modo en que aprenden los humanos¹. En otras palabras, el ML permite a las computadoras aprender de los datos y mejorar con la experiencia. Aquí tienes algunos puntos clave:

1. \*\***Aprendizaje automático**\*\*: El ML se basa en la idea de que las máquinas pueden aprender patrones y reglas a partir de datos sin ser programadas explícitamente. En lugar de seguir instrucciones específicas, las máquinas pueden adaptarse y mejorar su rendimiento a medida que se les proporciona más información.

2. \*\***Algoritmos y modelos**\*\*: En el ML, utilizamos algoritmos y modelos para procesar datos y extraer conocimientos. Estos algoritmos pueden clasificar, predecir, agrupar y descubrir patrones en los datos.

3. \*\***Tipos de ML**\*\*: Hay varios tipos de ML, como el aprendizaje supervisado (donde se proporcionan ejemplos etiquetados), el aprendizaje no supervisado (donde se agrupan datos sin etiquetas) y el aprendizaje por refuerzo (donde un agente aprende a través de interacciones con un entorno).

4. \*\***Aplicaciones**\*\*: El ML se utiliza en diversas áreas, como la detección de fraudes, la recomendación de productos, la visión por computadora, el procesamiento del lenguaje natural y más.

En resumen, el machine learning es una herramienta poderosa que permite a las máquinas aprender y mejorar a partir de datos, lo que tiene aplicaciones prácticas en muchos campos

**¿Qué es el Deep Learning?**

El deep learning, o aprendizaje profundo, es una rama del aprendizaje automático (machine learning) que se basa en redes neuronales artificiales con múltiples capas. Estas redes neuronales están inspiradas en la estructura y funcionamiento del cerebro humano y están diseñadas para imitar la forma en que los humanos aprenden y procesan la información.

Aquí hay algunos puntos clave sobre el deep learning:

**Redes Neuronales Profundas**: Las redes neuronales profundas (deep neural networks) consisten en varias capas de neuronas (también llamadas nodos). Cada capa transforma los datos de entrada en una representación más abstracta y de mayor nivel, lo que permite a la red aprender características complejas.

**Capas**: Hay varios tipos de capas en una red neuronal profunda, incluyendo capas de entrada, capas ocultas y capas de salida. Las capas ocultas son las que contienen la mayoría de las neuronas y son las responsables del "profundidad" del modelo.

**Algoritmos de Entrenamiento:** El entrenamiento de redes neuronales profundas suele utilizar algoritmos como la retropropagación (backpropagation) y el descenso de gradiente (gradient descent). Estos algoritmos ajustan los pesos y sesgos de las conexiones neuronales para minimizar el error en las predicciones de la red.

**Grandes Volúmenes de Datos**: El deep learning es particularmente eficaz cuando se tiene acceso a grandes volúmenes de datos y potentes recursos computacionales, ya que las redes neuronales profundas pueden tener millones de parámetros que necesitan ser ajustados.

**Aplicaciones:** El deep learning ha tenido un impacto significativo en una amplia variedad de campos, incluyendo la visión por computadora (por ejemplo, reconocimiento de imágenes y detección de objetos), procesamiento del lenguaje natural (por ejemplo, traducción automática y análisis de sentimientos), juegos, conducción autónoma, y más.

**Popularidad y Avances:** En los últimos años, el deep learning ha ganado popularidad debido a los avances en hardware (como GPUs y TPUs), la disponibilidad de grandes conjuntos de datos, y el desarrollo de marcos de trabajo de software como TensorFlow, PyTorch y Keras.

En resumen, el deep learning es una poderosa técnica de aprendizaje automático que permite a las computadoras aprender de datos de manera similar a cómo lo hace el cerebro humano, lo que ha llevado a avances significativos en muchas áreas tecnológicas.

**Ética de datos**

**“Los datos son personas”**

¿Qué es la ética y su relación con la inteligencia artificial?

Para empezar, es esencial comprender la base de nuestra discusión: la ética. Esta reflexión sobre la moral se empapa de nuestras creencias y juicios morales y analiza la conducta humana. En el contexto de la inteligencia artificial, la ética se convierte en un faro que guía el adecuado uso y manejo de los datos.

**¿Cómo definimos la moral en este contexto?**

Enlazado con la ética, la moral se compone de normas que regulan el comportamiento social humano, dictadas por el contexto cultural y legal. Estas normas son esenciales para poder convivir en una sociedad ordenada y justa.

**¿Qué importancia tiene la ética en el manejo de datos?**

La ética en el manejo de datos se centra en el uso responsable y correcto de la información que procesamos. Esto es crucial para todos los actores involucrados: usuarios, generadores de contenido y expertos en el tratamiento de datos.

**Principios claves de la ética en los datos**

Para profundizar, te presento cuatro preguntas esenciales que deben guiar nuestro análisis ético en la recolección y uso de datos:

* ¿Quién está recolectando la información? El origen de los datos es crucial — ya sea una organización pública o privada — pues define responsabilidades y estándares.
* ¿Por qué y para qué se recolectan los datos? Definir claramente el propósito del uso de la información y hacerlo de manera transparente es un pilar ético.
* ¿Cómo se recolectan y almacenan los datos? La seguridad y la mejora continua de las prácticas de recolección son vitales para la protección de los usuarios.
* ¿Cuál es el impacto de los análisis generados a partir de los datos? Reflexionar sobre el destino de la información y cómo favorece al bien común es indispensable.

**¿Por qué los datos son personas?**

Es importante tener presente que detrás de cada dato hay una persona, y con ello, la necesidad de cuidado y responsabilidad hacia la privacidad e identidad de cada individuo.

**Pilares de la Ethic & Datos**

Ahondaremos en los tres pilares fundamentales de la Ethic & Datos y las implicaciones prácticas para los profesionales:

**¿Cuál es el papel de la privacidad y la transparencia?**

*Privacidad*: Proteger la información personal de los usuarios es un deber primordial.

*Transparencia*: Comunicar de forma clara y accesible qué datos se recopilan y para qué fines.

**¿Cómo influye el diseño responsable de algoritmos?**

*Diseño Ético*: Es esencial diseñar algoritmos que sean inclusivos y eviten sesgos discriminatorios.

*Validación Ética:* La colaboración con expertos en ética para validar las prácticas tecnológicas es vital.

**¿Qué consideraciones se deben tener en las aplicaciones de los datos?**

Las aplicaciones tecnológicas deben seguir un código deontológico que garantice:

El consentimiento informado del usuario.

El uso ético de la tecnología.

Y la protección de la privacidad e identidad del usuario.

**¿Qué es Data4Good?**

Data4Good es un término que se utiliza para referirse a la utilización de datos con fines sociales y de beneficio público. Se trata de un enfoque que utiliza datos y tecnología para resolver problemas sociales, ambientales y económicos, y mejorar la vida de las personas.

**La importancia del Data4Good**

El Data4Good es una herramienta poderosa que puede ayudar a abordar algunos de los mayores desafíos que enfrenta la sociedad actualmente. A través de la recopilación, análisis y uso de datos, se pueden encontrar soluciones innovadoras a problemas complejos, como la pobreza, el cambio climático, la desigualdad, la salud y la educación.

**Ejemplos de Data4Good**

Algunos ejemplos de proyectos de Data4Good incluyen:

* El análisis de datos climáticos para predecir y mitigar los efectos del cambio climático.
* La utilización de la inteligencia artificial y el aprendizaje automático para mejorar la atención médica y prevenir enfermedades.
* El uso de datos para mejorar la calidad de la educación y la toma de decisiones educativas.
* La aplicación de datos para prevenir la violencia y el crimen en las ciudades.

**Ética en Data4Good**

Es importante destacar que, aunque el Data4Good tiene como objetivo mejorar la vida de las personas, también puede plantear desafíos éticos y de privacidad de los datos. Por esta razón, es fundamental que cualquier proyecto de Data4Good se realice con transparencia, responsabilidad y respeto a los derechos humanos y la privacidad de los datos.

**Conclusión**

En resumen, Data4Good es una herramienta poderosa para abordar problemas sociales y mejorar la vida de las personas. Sin embargo, es importante tener en cuenta que cualquier proyecto de Data4Good debe abordar los desafíos éticos y de privacidad de los datos para garantizar que se utilice de manera responsable y con el respeto adecuado por las personas y sus derechos.

**Introducción a Machine Learning**

**Objetivos:**

* Aprender a preparar datos y visualizarlos
* Entender diferentes algoritmos de machine learning
* Explorar Deep Learning y redes neuronales

**Data Science**

**Terminoligia para ciencia de datos**

**Datos/Datas:** Unidades de información o hechos de observaciones

**Features:** Tipos de información acerca de tus observaciones

**Filas:** Observaciones individuales o muestras.

**Columnas:** Features que describen tus observaciones

**Outlier:** Puntos de datos o data points que se comportan de forma extrañas

**Pre-Processing:**  Preparar datos para su uso en un modelo de machine learning

**ETL pipeline:**  Framework de data science para extraer, transformar y cargar

**Tipos de Data**

**Numéricos**: Su feature es un numero de tipo flotante

**Categórica**: Sus features representan una clase o tipo, usualmente se representan como un mapeo de números o un one-hot vector

**Imagen**: representa una imagen

**Texto**: Es la forma de texto sea corto (como twitter) o largo como noticias y artículos

**NaN**: Desconocido o corrupto o perdido o algo que no pertenece en el data set.

**Pandas**

esta librería es la ideal para cargar y entender tus datos. los comandos mas basicos son: pd.read\_csv(): Leer un archivo CSV df.head(): Muestra las primeras 5 filas df.dtypes: Muestra el tipo de representación de los datos (float, int, object).

**Visualizando tus datos**

los gráficos más famosos para analizar tus datos son: Histogramas: Te dice qué tan “frecuentes” y distribuidos son ciertos valores en tus datos. Gráficas de dispersión: Muestra la relación entre 2 features graficándolos como pares ordenados. Te puede ayudar a detectar anomalías.

**¿Qué es machine learning?**

El \*\*machine learning\*\* (ML) es una rama de la inteligencia artificial (IA) y la informática que se centra en el uso de datos y algoritmos para que la IA imite el modo en que aprenden los humanos¹. En otras palabras, el ML permite a las computadoras aprender de los datos y mejorar con la experiencia. Aquí tienes algunos puntos clave:

1. \*\*Aprendizaje automático\*\*: El ML se basa en el concepto de aprendizaje automático. Los algoritmos de ML pueden ajustarse automáticamente a los datos sin ser programados explícitamente. Esto les permite adaptarse y mejorar su rendimiento con el tiempo.

2. \*\*Tipos de algoritmos\*\*: Existen varios tipos de algoritmos de ML, como el aprendizaje supervisado (donde se utilizan datos etiquetados para entrenar modelos), el aprendizaje no supervisado (que busca patrones en datos no etiquetados) y el aprendizaje por refuerzo (donde un agente aprende a través de interacciones con un entorno).

3. \*\*Aplicaciones\*\*: El ML se utiliza en diversas áreas, como la detección de fraudes, la recomendación de productos, la visión por computadora, el procesamiento del lenguaje natural y más. Por ejemplo, los sistemas de recomendación de Netflix o las aplicaciones de reconocimiento facial utilizan técnicas de ML.

En resumen, el machine learning es una herramienta poderosa que permite a las máquinas aprender y mejorar a partir de datos, lo que tiene aplicaciones prácticas en muchos campos.

**Aprendizaje Supervisado**

El modelo obtiene features de entrada y salida, hay un Target/Objeto a predecir

**Regresión**: Target output (objeto de salida) es numérico

**Clasificación**: Target output es una etiqueta

**Aprendizaje no supervisado**

Objetivo desconocido, queremos encontrar estructura y grupos dentro de los datos

**Cloustering**: Queremos encontrar grupos en los datos

**Dimensionality reduction**: Queremos encontrar que features de entrada en los datos son de ayuda.

**Tipos de Machine Learning**

**Método Supervisado**

El modelo obtiene features de entrada y salida, hay un Target/Objeto a predecir

**Regresión**: Target output (objeto de salida) es numérico

**Clasificación**: Target output es una etiqueta

**Método no Supervisado**

Objetivo desconocido, queremos encontrar estructura y grupos dentro de los datos

**Cloustering**: Queremos encontrar grupos en los datos

**Dimensionality reduction**: Queremos encontrar que features de entrada en los datos son de ayuda.

**Data Scientist**

**¿Qué es la visualización de datos?**

1. **El input:** que es donde se encuentran los datos ya sea que estén organizados o no.
2. **Output:** es la presentación de los datos de una manera visual para ayudar a retener la información.

La visualización de datos sirve para reconocer patrones o encontrar más información para investigar.

Los mapas, los ábacos, son ejemplos de visualización de datos.

Utilizando la visualización de datos podemos organizar y mejorar el aprendizaje de datos en la era de la información.

**¿Qué es el Data Storytelling?**

El data storytelling combina datos con elementos narrativos para comunicar información de manera efectiva. En lugar de presentar solo números y gráficos, se crea una narrativa que ayuda a la audiencia a entender y recordar mejor la información.

**Elementos Clave del Data Storytelling**

1. **Narrativa**: Se estructura la información con un inicio, desarrollo y conclusión, similar a una historia tradicional.
2. **Visualización de Datos**: Se utilizan gráficos y visualizaciones para complementar la narrativa y hacer los datos más accesibles.
3. **Contexto**: Se proporciona el contexto necesario para que los datos tengan sentido y relevancia para la audiencia.
4. **Personajes y Emociones**: Se pueden incluir personajes o elementos emocionales para crear una conexión más fuerte con la audiencia.

**Beneficios del Data Storytelling**

* **Mejora la Comprensión**: Facilita la comprensión de datos complejos.
* **Aumenta la Retención**: Las historias son más fáciles de recordar que los datos aislados.
* **Genera Impacto**: Puede influir en la toma de decisiones al presentar datos de manera convincente.

**¿Qué son los KPIs?**

Los **KPIs** (Key Performance Indicators) o **Indicadores Clave de Rendimiento** son métricas cuantitativas que se utilizan para medir el progreso hacia objetivos específicos dentro de una organización. Aquí tienes un resumen:

Los KPIs son herramientas que ayudan a las empresas a evaluar su éxito en alcanzar metas estratégicas. Estos indicadores proporcionan una forma clara y objetiva de medir el rendimiento en diversas áreas.

**Características de los KPIs**

1. **Específicos y Medibles**: Deben ser claros y cuantificables.
2. **Relevantes**: Deben estar alineados con los objetivos estratégicos de la organización.
3. **Temporales**: Deben tener un marco de tiempo definido para su evaluación.

**Tipos de KPIs**

* **KPIs Financieros**: Como el retorno sobre la inversión (ROI) y el margen de beneficio.
* **KPIs de Clientes**: Como la satisfacción del cliente y la tasa de retención.
* **KPIs de Procesos Internos**: Como la eficiencia operativa y la calidad del producto.
* **KPIs de Crecimiento y Aprendizaje**: Como la tasa de capacitación y el desarrollo de habilidades.

**Beneficios de Utilizar KPIs**

* **Monitoreo del Progreso**: Permiten seguir el avance hacia los objetivos.
* **Toma de Decisiones**: Facilitan la identificación de áreas que necesitan mejoras.
* **Motivación**: Ayudan a mantener a los equipos enfocados y motivados.

**Ejemplos de KPIs**

* **Ventas Mensuales**: Medir el total de ventas realizadas en un mes.
* **Tasa de Conversión**: Porcentaje de visitantes que realizan una acción deseada (como una compra).
* **Tiempo de Resolución de Problemas**: Tiempo promedio que se tarda en resolver un problema de cliente