



Evaluación - Científico de Datos

Conocimientos generales:

1. Explicar brevemente la definición de Inteligencia Artificial y fundamentar su opinión y posibilidades de aplicación actuales y a futuro de esta.

La Inteligencia Artificial es una rama de la informática que estudia y desarrolla sistemas capaces de percibir su entorno, aprender de los datos y tomar decisiones racionales orientadas a objetivos. Actualmente, sus aplicaciones incluyen el diagnóstico médico mediante modelos de aprendizaje supervisado, la detección de fraude financiero con análisis predictivo, el procesamiento de lenguaje natural con modelos de gran escala y los sistemas de recomendación basados en filtrado colaborativo.

En mi opinión, nos encontramos en un punto de inflexión tecnológico, donde la IA está dejando de ser experimental para convertirse en infraestructura crítica. A futuro, su impacto será mayor en agentes autónomos, automatización cognitiva y sistemas híbridos humano-máquina. Sin embargo, los desafíos principales no son únicamente técnicos, sino también éticos y regulatorios, especialmente en términos de sesgo, gobernanza y equidad.

2. Explicar brevemente la diferencia entre la definición de **Strong AI** y **Weak AI**.

Weak AI (o Narrow AI) es lo que existe hoy: sistemas diseñados para una tarea específica, como GPT para texto o AlphaGo para ajedrez. Son muy buenos en su dominio pero no generalizan fuera de él. Strong AI (o AGI) es un sistema con capacidad cognitiva general comparable a la humana, capaz de razonar, aprender y adaptarse a cualquier tarea. Hoy todo lo que usamos en producción es Weak AI. Strong AI sigue siendo un objetivo teórico sin implementación real.

3. Explicar las diferencias encontradas con base a su experiencia entre **Machine Learning**, **Deep Learning** e **Inteligencia Artificial**.

Se relacionan como círculos concéntricos: IA es el campo más amplio, ML es un subconjunto, y DL es un subconjunto de ML. IA incluye cualquier técnica que haga a una máquina comportarse de forma inteligente, incluso sistemas basados en reglas. ML es cuando la máquina aprende patrones directamente desde los datos sin ser programada explícitamente. Deep Learning usa redes neuronales profundas y destaca con datos no estructurados como imágenes, audio o texto. En la práctica, para datos tabulares estructurados sigo prefiriendo ML clásico (XGBoost, Random Forest) por su interpretabilidad y eficiencia; DL lo reservo para problemas de visión, NLP o patrones no triviales.

4. Explicar qué es una ANN. ¿Cuál es su definición? ¿Cómo se conocen sus diferentes capas?

Una ANN (Artificial Neural Network) es un modelo computacional inspirado en el cerebro humano, compuesto por nodos organizados en capas que procesan información mediante pesos y funciones de activación. Sus capas se clasifican en tres tipos: la capa de entrada, que recibe los datos en crudo; las capas ocultas, que aprenden representaciones intermedias y donde ocurre la mayor parte del aprendizaje; y la capa de salida, que produce la predicción final. Entre más capas ocultas tenga la red, más profunda es.

5. Con base en su experiencia, ¿Cómo se relaciona la Teoría de Juegos con la Inteligencia Artificial? Esta pregunta me resultó interesante.

La Teoría de Juegos estudia la toma de decisiones estratégica entre agentes racionales, y su conexión con la IA es directa y profunda. El ejemplo más claro son las GANs (redes adversarias generativas), donde un generador y un discriminador juegan un juego de suma cero hasta alcanzar un equilibrio de Nash (otro concepto interesante que vi en la universidad). También aparece en Reinforcement Learning multiagente, donde varios agentes aprenden políticas óptimas considerando las acciones de los demás. Quizás existan otros ejemplos pero por el momento estos describen mi experiencia.

6. ¿A través de qué modelo y técnica computacional solucionaría un problema de complejidad NP-duro?

Los problemas NP-duros no tienen solución exacta eficiente en tiempo polinomial, así que la estrategia es buscar soluciones aproximadas de buena calidad. Yo combinaría dos enfoques: metaheurísticas como Algoritmos Genéticos o evolutivos para explorar el espacio de soluciones, y Reinforcement Learning para que el agente aprenda heurísticas de solución a través de experiencia.

La solución depende mucho del modelado en realidad y la comprensión del problema.

7. Explica los tipos de EDA.

El análisis exploratorio de datos (EDA) puede clasificarse según la cantidad de variables que se analizan. El EDA univariado estudia una variable de manera aislada, con el fin de entender su distribución, tendencia central y dispersión, así como detectar posibles outliers. Para ello se emplean medidas como la media, mediana, varianza y desviación estándar, además de representaciones gráficas como histogramas o diagramas de caja.

El EDA bivariado examina la relación entre dos variables, buscando identificar asociaciones, correlaciones o dependencias. Dependiendo del tipo de variables (cuantitativas o categóricas), se utilizan herramientas como coeficientes de correlación, tablas de contingencia o gráficos de dispersión. Finalmente, el EDA multivariado analiza simultáneamente más de dos variables, permitiendo detectar patrones estructurales más complejos mediante matrices de correlación, mapas de calor o técnicas de reducción de dimensionalidad.

8. ¿Qué es una prueba A/B y cómo se analizan los resultados?

Una prueba A/B es un experimento controlado donde se divide aleatoriamente a los usuarios en dos grupos: el grupo de control (A) que recibe la experiencia actual, y el grupo de tratamiento (B) que recibe la nueva versión. El objetivo es medir si el cambio produce una diferencia estadísticamente significativa en una métrica clave. Para analizar los resultados primero defino la hipótesis nula (no hay diferencia entre A y B) y elijo una métrica principal. Luego aplico una prueba estadística adecuada: t-test para medias continuas, prueba chi-cuadrado para proporciones o tasas de conversión. Si el p-valor es menor al umbral definido (usualmente 0.05) y el tamaño del efecto es práctico, rechazo la hipótesis nula y concluyo que el cambio tiene impacto real.

9. Explica el Compensación Bias - Varianza

El error total de un modelo se puede descomponer en tres partes: bias, varianza y ruido irreducible. El bias (sesgo) mide qué tan lejos están las predicciones del modelo del valor real en promedio; un bias alto significa que el modelo es demasiado simple y no captura la estructura de los datos (underfitting). La varianza mide qué tanto cambian las predicciones del modelo al entrenarlo con distintos subconjuntos de datos; una varianza alta significa que el modelo memoriza el ruido del entrenamiento y no generaliza (overfitting). El tradeoff es que reducir uno tiende a aumentar el otro. La solución práctica está en encontrar el punto de equilibrio.

Análisis

De acuerdo con el siguiente set de datos que corresponde a los datos de las viviendas de California, realice un **análisis exploratorio**, identifique **datos atípicos**, realice **limpieza y proponga dos modelos que predigan el costo de las viviendas**. Y finalmente, elabore un **reporte operativo** y otro **directivo** con los hallazgos identificados.

Layout

1. longitude: Una medida de cuán al oeste está una casa
2. latitude: Una medida de cuán al norte está una casa
3. housingMedianAge: Edad media de una casa dentro de una cuadra; un número más bajo indica un edificio más nuevo.
4. totalRooms: Número total de habitaciones dentro de una cuadra.
5. totalBedrooms: Número total de dormitorios dentro de una cuadra.
6. population: Número total de personas que viven en una cuadra.

Documentos anexados y entregados.

Se pueden consultar también en : https://github.com/Juan-Alvarado21/california_housing_rep-analysis