

Principios de Big Data

Clase 1 — ¿Qué es y por qué está cambiando la medicina?

Cristian Lazo Quispe

Diplomado en Ciencia de Datos en Salud · Universidad Científica del Sur

17 de junio de 2026

Curso **Principios de Big Data** · Diplomado en Ciencia de Datos en Salud (DICDASA)
Universidad Científica del Sur · Docente responsable del diplomado: Mg. Jorge L. Maguiña.

✉ clazoq@uni.pe ·  cristian-lazo-quispe ·  CristianLazoQuispe

Itinerario de hoy (6:00 – 10:00 pm)

6:00 – 6:15	Bienvenida y por qué esto te importa
6:15 – 7:30	Bloque 1: Las palabras (Big Data, IA, ML, Deep Learning, datos)
7:30 – 7:45	<i>Pausa activa</i>
7:45 – 8:45	Bloque 2: Ejemplos con imágenes + herramientas + errores comunes
8:45 – 9:00	<i>Pausa activa</i>
9:00 – 9:40	Demo en vivo: “siente la escala” de los datos
9:40 – 10:00	Cómo se decide un proyecto + cierre

Para llevarte

Hoy no vienes a programar. Vienes a **entender, juzgar y decidir.**

¿Qué veremos hoy?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala
- 9 Cómo se decide un proyecto
- 10 Cierre

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

Tres historias reales (no de Silicon Valley)

Esto ya está pasando en hospitales como el tuyo

- Un modelo marcó qué pacientes con **insuficiencia renal** tenían alta probabilidad de **reingresar en 30 días** $\Rightarrow$ el equipo priorizó seguimiento y **evitó reingresos**.
- Un algoritmo ayudó a **detectar lesiones** en imágenes de baja calidad de zonas rurales, donde no hay especialista disponible.
- Analizando datos de un hospital, los pacientes pediátricos con **dengue** que llegaban ciertos días tenían peor evolución — no por la enfermedad, sino por **cómo estaba organizada la guardia**.

Para llevarte

El dato no reemplaza al médico. Le da **una segunda mirada** y **anticipa**.

Qué te llevarás de este curso (4 clases)

Vas a poder...

- Entender de qué se habla cuando dicen “Big Data / IA”.
- **Juzgar** si una propuesta tecnológica es seria o “humo”.
- **Encuadrar** un proyecto de datos en tu servicio.
- Saber **a quién contratar** y qué preguntarle.

NO necesitas...

- Programar ni escribir código.
- Saber matemáticas avanzadas.
- Instalar nada en tu computadora.

Lo único que tocarás: **una caja de texto** para hacerle preguntas a los datos.

Cuéntanos, en una frase

- Tu **nombre** y tu **especialidad** / **área**.
- **Un momento** en que la tecnología o los datos te **ayudaron** — o te **complicaron** — en tu trabajo.

Para llevarte

No hay respuestas técnicas. Quiero conocer **tu día a día**: de ahí saldrán los ejemplos de hoy.

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

Usamos estas palabras como sinónimos (y no lo son)

Palabras que se confunden todo el tiempo

Datos · Big Data · Inteligencia Artificial · Machine Learning
Deep Learning · Automatización · Ciencia de Datos · “un dashboard”

Para llevarte

Si no distingues estas palabras, es fácil que te **vendan humo**. Hoy las ordenamos, con ejemplos.

¿Qué es Big Data? (en una frase)

Big Data

Técnicas y tecnologías para **capturar, guardar y analizar** datos **tan grandes o tan rápidos** que las herramientas de siempre (una hoja de Excel) **ya no alcanzan**.

Analogía clínica

Una cosa es revisar la historia de *un* paciente. Otra es mirar, a la vez, la de **2 millones** de pacientes para encontrar el **patrón** que a simple vista no se ve.

SÍ es Big Data

- Señales de **todos** los monitores de UCI, en tiempo real.
- **Todas** las imágenes (rayos X, TC) del hospital, por años.
- Registros de millones de atenciones del **SIS / MINSA**.

NO es Big Data

- Una planilla de **200 pacientes** en Excel.
- El registro mensual de tu consultorio.
- Un PDF con la lista de turnos.

Eso son **datos** —valiosos, pero los maneja una persona.

Tu turno: ¿es Big Data o no? (a mano alzada)

Levanta la mano en cada caso

Voy diciendo cada situación y votas con la mano: ¿es Big Data o no?

- 1 Señales de **todos los monitores de UCI**, en tiempo real.
- 2 Una **planilla de 200 pacientes** en Excel.
- 3 El **archivo de radiología** del hospital, de varios años.
- 4 El **registro mensual** de tu consultorio.

Para llevarte

No hay nota ni respuesta perfecta: lo que importa es **por qué** votaste así. Sirve para **pensar juntos** dónde empieza el “Big Data”.

La pregunta del millón: ¿cuántos datos hacen falta?

No hay un número mágico

No es “a partir de 1 millón de filas”. Big Data es **relativo**: empieza cuando los datos **superan lo que una sola computadora / herramienta común** puede manejar con comodidad.

Reglas prácticas (si se cumple alguna, piensa en Big Data)

- No cabe en Excel ni en la memoria de una PC.
- Tarda demasiado en procesarse en una sola máquina.
- **Crece más rápido** de lo que puedes manejar (llega solo, en tiempo real).
- Es muy **variado** (texto + imágenes + señales) y no entra en una tabla.

Checklist: ¿cuándo llamar “Big Data” a un sistema?

Las 5 V (criterio clásico):

- **Volumen** grande **Velocidad** alta
- **Variedad** de formatos **Veracidad**
- **Valor** para una decisión

+ **señal técnica**: necesita **procesamiento distribuido** (varias máquinas trabajando juntas).

Variables a definir (del enfoque clásico):

- Volumetría (¿cuánto?)
- Tipo de datos (¿estructurado o no?)
- Nivel de seguridad
- Ubicación: **on-premise vs nube**
- Disponibilidad (¿en tiempo real?)

Para llevarte

¿Necesitas nube y base de datos? **No siempre**, pero a gran escala ayudan mucho. El dato puede ser estructurado o no.

¿Cuándo SÍ y cuándo NO conviene usar Big Data?

SÍ conviene

- Mucho volumen / tiempo real (UCI, wearables).
- Buscar patrones en **poblaciones**.
- Imágenes, texto libre, señales a gran escala.

NO conviene (no te endeudes)

- Tu pregunta cabe en Excel.
- Problema pequeño o puntual.
- El dato es de mala calidad o no existe.

Para llevarte

Big Data es un **medio**, no un trofeo. Úsalo cuando el problema lo pide.

No hay un umbral oficial de filas

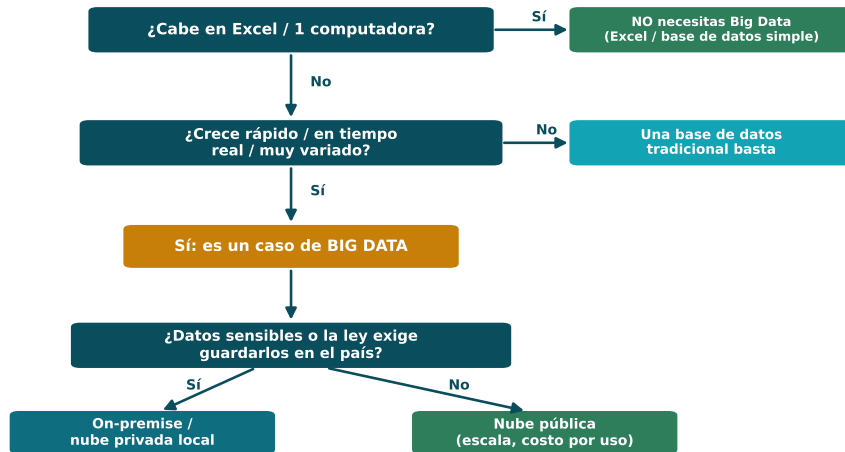
Ninguna definición sería decir “Big Data = a partir de X datos”. Las definiciones reconocidas se basan en **características**, no en un número:

- **Gartner / Doug Laney (2001)**: las **3 V** — Volumen, Velocidad, Variedad.
- **NIST** (*Big Data Interoperability Framework*): es Big Data cuando el volumen/velocidad/variedad **obliga** a usar **arquitecturas distribuidas** (varias máquinas) para procesarlo.

Para llevarte

Regla práctica: **si una sola computadora ya no puede con comodidad**, entras en Big Data. El “número” depende de tu máquina y tu problema, no es fijo.

Diagrama de decisión: ¿Big Data? ¿on-premise o nube?



On-premise vs. nube: ¿cuándo cada una?

	On-premise (en tu hospital)	Nube (cloud)
Control / datos	Máximo control; útil si la ley exige guardarlos en el país	El proveedor opera la infraestructura
Costo	Gran inversión inicial (servidores)	Pago por uso; sin servidores propios
Escala	Limitada por tu hardware	Escala casi ilimitada y rápida
Ideal para	Datos muy sensibles, normativa estricta	Crecer rápido, picos, proyectos nuevos

Para llevarte

Muchos hospitales usan un **modelo híbrido**: lo sensible on-premise, el análisis a gran escala en la nube.

Las “V” de la Big Data (en términos de salud)

Volumen	Millones de registros: historias clínicas, imágenes, laboratorio.
Velocidad	Datos en tiempo real: un monitor en UCI, un wearable.
Variedad	Texto, imágenes (rayos X), señales (ECG), genómica, audio.
Veracidad	¿El dato es confiable? Errores de digitación, equipos mal calibrados.
Valor	Lo único que importa al final: ¿mejora una decisión clínica?

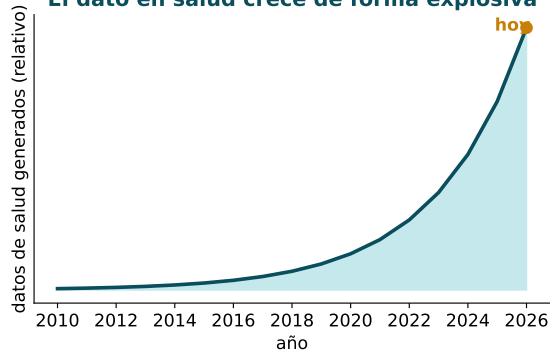
Para llevarte

Las dos que más te deben preocupar como líder son **Veracidad** y **Valor**.

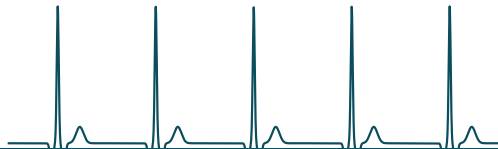
¿Cuánto dato genera la medicina hoy?

- **Monitor cardíaco:** ~86 400 lecturas por paciente al día.
- **Tomografía:** cientos de imágenes por estudio.
- **Wearable:** ritmo, sueño, oxígeno, 24/7.
- **Genoma:** varios *gigabytes* por paciente.

El dato en salud crece de forma explosiva



Una señal (ECG): el dato llega sin parar, 24/7



Dos tipos de dato: estructurado vs. no estructurado

ESTRUCTURADO

Edad	PA	Glucosa
67	140/90	126
54	120/80	98

NO ESTRUCTURADO

“paciente refiere dolor torácico...”

imagen (rayos X)

señal de ECG

genoma

Para llevarte

El **~80 %** del dato en salud es **no estructurado** (notas, imágenes): el más valioso y el más difícil de aprovechar.

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning**
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

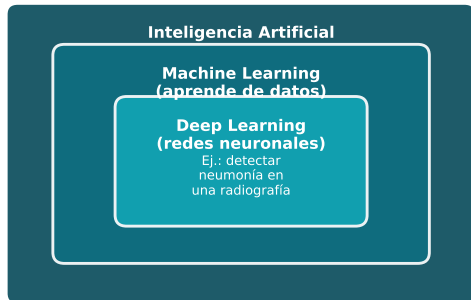
¿Qué es IA? ¿Y qué NO es IA?



- **IA** (que aprende): detecta neumonía en una placa nueva.
- **Automatización**: “si fiebre > 38, alerta” (regla fija).

Para llevarte

La IA **aprende patrones** de los datos; la automatización **sigue reglas** que alguien escribió.



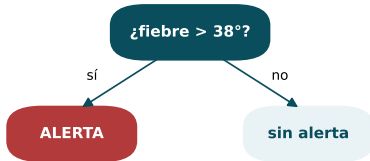
- **IA:** la idea grande (máquinas que “deciden”).
- **Machine Learning (ML):** IA que **aprende de ejemplos**.
- **Deep Learning:** ML con **redes neuronales**; brilla en imágenes/texto.

IA vs ML: toda ML es IA, pero **no** toda IA es ML (hay IA basada en reglas, sin aprender).

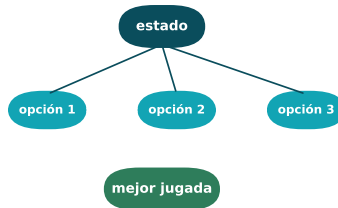
5 ejemplos: Inteligencia Artificial SIN Machine Learning

- 1 **Sistema experto** tipo MYCIN: recomienda antibiótico según reglas escritas.
- 2 **Verificador de interacciones** entre medicamentos (base de reglas).
- 3 **Escalas clínicas automatizadas** (Wells, CURB-65): aplican un criterio fijo.
- 4 **Ajedrez clásico** (Deep Blue): busca jugadas, no “aprende”.
- 5 **Ruta más corta / GPS** (algoritmo de Dijkstra) o **chatbot de árbol fijo**.

Regla fija (si... entonces...)



Búsqueda (ej.: ajedrez/ruta)

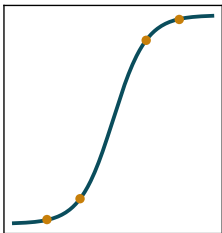


Son “inteligentes” porque resuelven, pero un humano programó la regla.

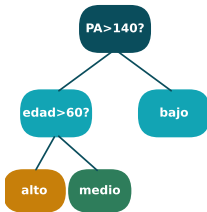
5 ejemplos: Machine Learning SIN Deep Learning

- 1 **Regresión logística:** riesgo cardiovascular (estudio de Framingham).
- 2 **Random forest / árboles:** predicción de reingreso hospitalario.
- 3 **k-means:** segmentar pacientes en grupos parecidos.
- 4 **Naïve Bayes:** clasificar texto (p. ej. correos o notas).
- 5 **Gradient boosting (XGBoost):** alerta temprana de sepsis.

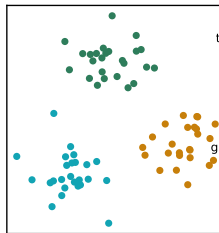
Regresión logística
(riesgo)



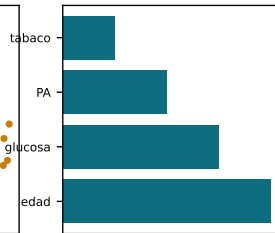
Árbol de decisión
(reingreso)



k-means
(segmentar)



Importancia
de variables



Aprenden de ejemplos y son explicables: puedes ver qué variable pesó más.

5 ejemplos: Deep Learning (redes neuronales)

Brilla con imágenes, señales y texto

- ① **Neumonía** en radiografía de tórax (visión por computadora).
- ② **Retinopatía diabética** en fondo de ojo.
- ③ **Cáncer de piel** (melanoma) a partir de fotos.
- ④ **Arritmias** en la señal de ECG.
- ⑤ **Modelos de lenguaje (LLM)**: resumir la historia clínica, responder preguntas.

Para llevarte

Muy potente, pero más **“caja negra”**: cuesta explicar *por qué* decidió.

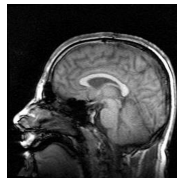
Galería: la IA “leyendo” imágenes médicas



Radiografía de tórax



TC de tórax



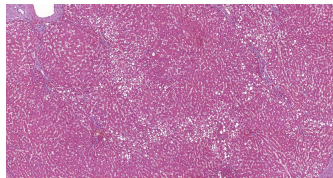
RM cerebral



Fondo de ojo



Lesión de piel



Patología (microscopio)

La misma técnica (Deep Learning) aprende a señalar hallazgos en cada tipo de imagen.

Ciencia de Datos: ¿qué es y qué NO es?



Ciencia de Datos = todo el camino, no solo el modelo

SÍ es Ciencia de Datos

Todo el camino: limpiar el dato, analizarlo, modelar y **convertirlo en decisión**.

NO es Ciencia de Datos

Solo hacer un gráfico bonito, o “tener” los datos guardados sin usarlos.

El mapa completo (para no confundirte nunca más)

- **Big Data** — el *material*: muchísimos datos.
- **Ciencia de Datos** — el *oficio*: sacar conocimiento (estadística, gráficos, ML).
- **Inteligencia Artificial** — el *objetivo*: que la máquina decida o prediga.
- **Machine / Deep Learning** — el *cómo* aprende la IA (de ejemplos / con redes).
- **Automatización** — tareas con *reglas fijas* (no siempre es IA).
- **Dashboard / BI** — *mostrar* datos (no es “inteligencia”).

Para llevarte

Big Data = la harina. Ciencia de Datos = el panadero. IA = el pan que decide solo. Dashboard = la vitrina.

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa**
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

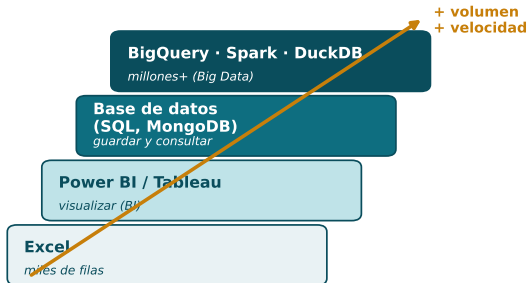
Excel, Power BI, base de datos, BigQuery, MongoDB... ¿qué es qué?

Herramienta	Para qué sirve	Cuándo usarla
Excel	Tablas pequeñas, cálculos a mano	Cientos/miles de filas
Power BI / Tableau	Visualizar (dashboards, BI)	Mostrar indicadores claros
Base de datos SQL (PostgreSQL)	Guardar y consultar datos ordenados	Muchos registros estructurados
MongoDB	Guardar datos no estructurados (documentos)	Texto, formularios variables
BigQuery · Spark · DuckDB	Analizar a gran escala (Big Data)	Millones de filas, en la nube

Para llevarte

“Power BI” **no** es IA: es una *vitrina* de datos. Confundirlos es el error #1.

¿Cuándo necesito cada herramienta? (intuición)

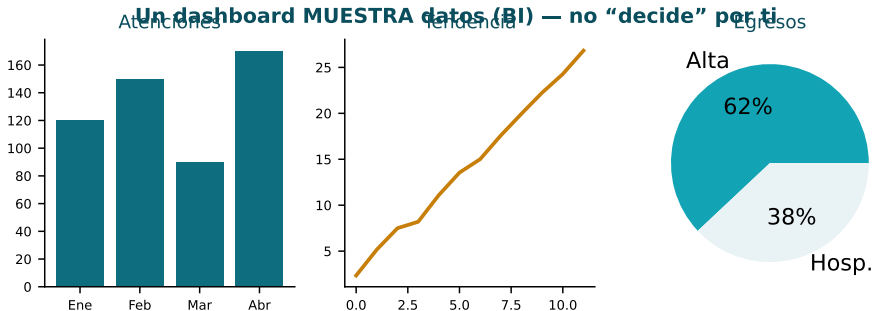


A medida que crecen el **volumen** y la **velocidad**, subes de escalón.

Para llevarte

No pagues Big Data si tu problema cabe en Excel. Y no fuerces Excel cuando ya necesitas una base de datos.

Un dashboard MUESTRA datos — no “decide”



Para llevarte

Útil para **ver** lo que pasa (Business Intelligence). Pero por sí solo **no predice ni recomienda**: eso es IA.

¿Cómo se ve una “base de datos” / dataset en salud?

paciente_id	edad	diagnóstico	HbA1c	región	reingreso_30d
P-0001	67	Diabetes tipo 2	8.1	Lima	Sí
P-0002	54	Hipertensión	—	Cusco	No
P-0003	72	Insuf. renal	7.4	Loreto	Sí
P-0004	45	EPOC	—	Arequipa	No
P-0005	60	Diabetes tipo 2	9.0	Piura	Sí

Datasets reales y abiertos para aprender (de referencia): **MIMIC** (cuidados intensivos), **Framingham** (riesgo cardiovascular), **eICU**, y los **datos abiertos del MINSA**.

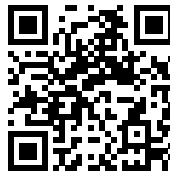
Para llevarte

Una fila = un caso; una columna = una variable. El **Big Data** aparece cuando esta tabla tiene *millones* de filas o se llena *so/a* en tiempo real.

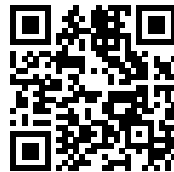
Exploremos datos públicos reales (sin instalar nada)

Dos fuentes abiertas que abriremos en vivo

- **Datos Abiertos del Perú** (MINSA y más): tablas reales del país.
- **Our World in Data**: gráficos interactivos de salud (la curva de COVID que vimos antes).



Datos Abiertos Perú



Our World in Data

Escanea para abrirlas en tu celular y “tocar” los datos mientras conversamos.

Para llevarte

Nota la diferencia: una **tabla** que descargas vs. datos que se actualizan **solos**.

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes**
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

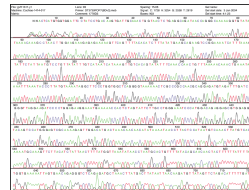
¿De dónde sale el Big Data en salud? (fuentes)



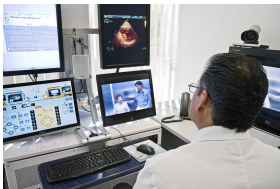
Wearables (24/7)



Monitor de UCI



Secuenciación genómica

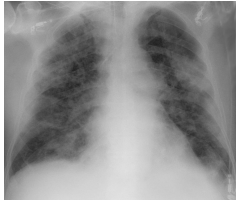


Historia clínica / telemedicina

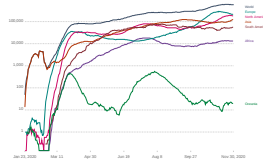


Centro de datos

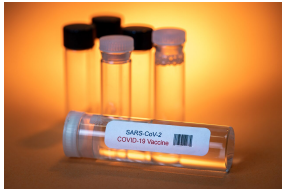
Big Data a escala poblacional



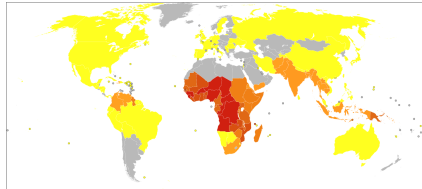
Rx de neumonía por COVID-19



Curva epidémica (casos por continente)



Cadena de frío de vacunas



Mapa de mortalidad (poblacional)

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes**
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

Mitos que un líder de salud debe saber desarmar

Lo que se dice	La realidad
"Tengo un Excel grande, eso es Big Data."	No: Big Data es cuando <i>ya no</i> cabe en Excel.
"Un dashboard bonito es IA."	No: es <i>mostrar</i> datos (BI), no decidir.
"La IA siempre acierta."	No: se equivoca, y con datos sesgados, más.
"Mientras más datos, mejor."	No: datos malos $\Rightarrow$ peores decisiones.
"Compro el software y listo."	No: sin adopción del personal, queda en un cajón.

Para llevarte

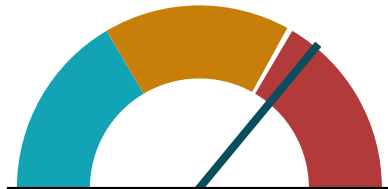
Tu superpoder como líder: hacer **la pregunta incómoda** antes de firmar el cheque.

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

Lo que SÍ puede hacer (capacidades reales)

- **Predecir riesgo:** reingreso, sepsis, no-asistencia a cita.
- **Apoyar el diagnóstico por imágenes:** radiología, patología.
- **Copilotos generativos:** redactar la nota, resumir la historia.
- **Salud poblacional:** dónde poner recursos, detectar brotes.



Predicción de riesgo (ejemplo).

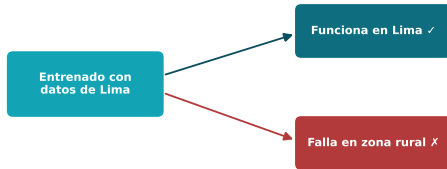
Para llevarte

Patrón común: **anticipar y ordenar el trabajo**, no “curar solo”.

Lo que NO puede (y dónde se equivoca)

Riesgos que un líder debe vigilar

- Basura entra, basura sale.
- Sesgo: entrena en una población, falla en otra.
- Caja negra: acierta pero no explica.
- Privacidad: datos sensibles reales.



Para llevarte

Tu trabajo no es confiar ciegamente: **pregunta** con qué datos se entrenó, en quién se validó y qué pasa cuando se equivoca.

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala**

Lo que vas a ver (sin que programes nada)

- ① Generamos datos de sensores de pacientes: **millones** de lecturas.
- ② Los abrimos con la herramienta “de siempre” (Excel / pandas): **sufre**.
- ③ La misma pregunta con una herramienta de Big Data (**DuckDB**): responde en **segundos**.

La frase para recordar

“Pandas y Excel se diseñaron cuando un hospital tenía miles de registros al año. Hoy **un solo monitor** genera más que eso en **un día**.”

(Notebook: demos/01_escala_bigdata.ipynb — solo presionar “Ejecutar”).

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

El Canvas de Alcance (tu herramienta del curso)

Antes de gastar un sol en tecnología, responde **6 preguntas**:

- 1 **Problema**: ¿qué decisión clínica/gestión quiero mejorar?
- 2 **Resultado medible**: ¿cómo sabré que funcionó? (un número)
- 3 **Datos**: ¿qué dato necesito y qué tan ordenado está?
- 4 **Equipo**: ¿a quién necesito contratar?
- 5 **Riesgos / ética**: sesgo, privacidad, ¿qué pasa si falla?
- 6 **Adopción**: ¿quién lo va a usar y cómo logro que lo use?

Para llevarte

El error #1 es saltar a la solución sin definir el problema. En la Clase 4 lo aplicarás a **un problema de tu propio hospital** (proyecto final “AI Charter”).

¿A quién contratar? (lenguaje de director)

Líder clínico / dueño	Define el problema y mide el valor. (Puedes ser tú.)
Ingeniero de datos	Ordena y limpia el dato (lo más caro y subestimado).
Científico de datos / ML	Construye el modelo.
Gestión del cambio	Logra que el personal realmente lo use.

Lo que más predice el éxito

No es el algoritmo: es la **adopción**. “Excelencia técnica sin uso = gasto en un cajón.”

¿Dónde estamos?

- 1 Por qué esto te importa
- 2 Pongámonos de acuerdo en las palabras
- 3 IA, Machine Learning y Deep Learning
- 4 Las herramientas: qué es cada cosa
- 5 Galería: el Big Data en salud, con imágenes
- 6 Errores y confusiones comunes
- 7 Qué SÍ y qué NO puede hacer
- 8 Demo en vivo: siente la escala

Hacia dónde vamos (las 4 clases)

Clase 1	Hoy	Qué es Big Data y por qué cambia la medicina.
Clase 2	Sáb 20 jun	IA en la práctica clínica + <i>preguntarle a la base de datos en español.</i>
Clase 3	Lun 6 jul	Datos, interoperabilidad e IA responsable + <i>chatbot sobre guías clínicas.</i>
Clase 4	Mié 8 jul	Liderar proyectos: tu “AI Charter” + segmentar pacientes.

Para llevarte

Sin tarea formal. Solo **una reflexión**: piensa en **un problema** de tu servicio que te gustaría resolver con datos. Será la semilla de tu proyecto final (Clase 4).

Gracias.

¿Qué problema de tu hospital resolverías con datos?

✉ clazoq@uni.pe ·  cristian-lazo-quispe ·  CristianLazoQuispe

Diplomado en Ciencia de Datos en Salud · Universidad Científica del Sur

Definiciones de Big Data

- Laney, D. (2001). *3D Data Management: Controlling Data Volume, Velocity, and Variety*. META Group / Gartner. (Origen de las “3 V”.)
- NIST (2019). *NIST Big Data Interoperability Framework*, Special Publication 1500 (vol. 1: Definiciones). National Institute of Standards and Technology, EE.UU.
- Mayer-Schönberger, V. & Cukier, K. (2013). *Big Data: A Revolution That Will Transform How We Live, Work, and Think*.

Datos e IA en salud

- Topol, E. (2019). *Deep Medicine: How Artificial Intelligence Can Make Healthcare Human Again*.
- OECD (2015). *Health Data Governance*. OECD Publishing.
- OMS / OPS — Estrategia mundial sobre salud digital 2020–2025.

Referencias para profundizar; no es necesario leerlas para el curso.

Imágenes médicas (Wikimedia Commons, uso educativo):

- Radiografía de tórax — James Heilman, MD — CC BY-SA 3.0.
- TC de tórax — Ptrump16 — CC BY-SA 4.0.
- RM cerebral — Oliver Stollmann — CC BY 3.0.
- Mamografía — U.S. National Cancer Institute / NIH — Dominio público.
- Lesión de piel (melanoma) — Simone Trovato Monastra — CC BY-SA 4.0.
- Fondo de ojo — Mikael Häggström — CC0 / Dominio público.
- Histología hepática — Oktay98 — CC BY-SA 4.0.

Dispositivos y fuentes de datos (Wikimedia Commons):

- Reloj inteligente — whity (Flickr) — CC BY 2.0.
- Monitor de signos vitales — PO1 J. Stenberg, U.S. Navy — Dominio público.
- Secuenciación de ADN — NHGRI — Dominio público.
- Centro de datos — Carl Lender — CC BY 2.0.
- Estación clínica / telemedicina — Intel Free Press — CC BY-SA 2.0.

Salud pública:

- Rx neumonía COVID-19 — Hellerhoff — CC BY-SA 3.0.
- Curva epidémica — Our World in Data — CC BY 4.0.
- Viales de vacuna — Cindy Shebley — CC BY 2.0.
- Mapa de mortalidad por malaria (OMS 2012) — Chris55 — CC BY-SA 4.0.

Gráficos propios: elaboración propia con Python/matplotlib.