



# Semana 24:

## Tipos de datos en Sanidad

---

*Jesús Silva Rodríguez, PhD*

# Estructura del módulo 3

## Clases:

Semana 24: Introducción: Tipos de datos en Sanidad.

Semana 25: Sistemas Informáticos: HIS, HCE

Semana 26: Imagen médica: PACS + DICOM

Semana 27: Natural Language Processing (Práctica)

Semana 28: Automatización Robótica de Procesos (Práctica)

Semana 29: Análisis de Imagen Médica + CDSS (Práctica)

Semana 30: Análisis de Imagen Médica + CDSS (Práctica)

Semana 31: Análisis de Imagen Médica + CDSS (Práctica)

Semana 32: Trabajo de Fin de módulo.

## Qué es un dato?

-



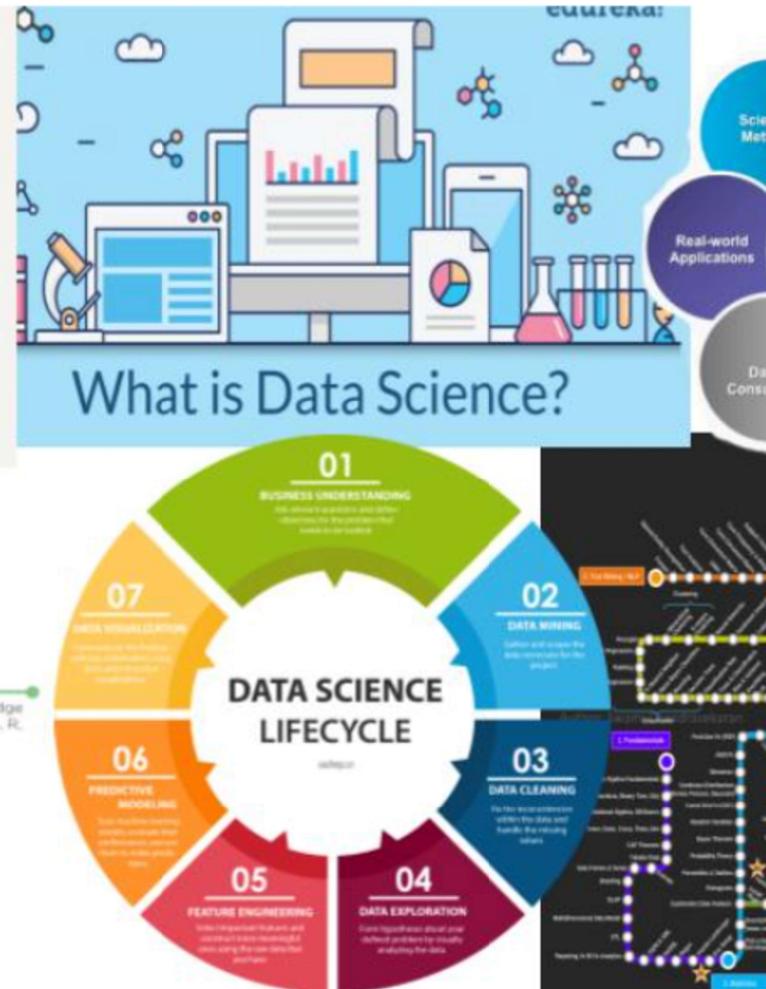
# Ciencia de datos

# Qué es la ciencia de datos?

---

- La **ciencia de datos** es un campo interdisciplinario que, con métodos científicos, extrae conocimiento de datos estructurados o no estructurados<sup>1</sup>. Es una combinación de campos de análisis de datos como la **estadística**, la **minería de datos**, el **aprendizaje automático**, y la **analítica predictiva**.<sup>1</sup>

gle!



# Big data

---



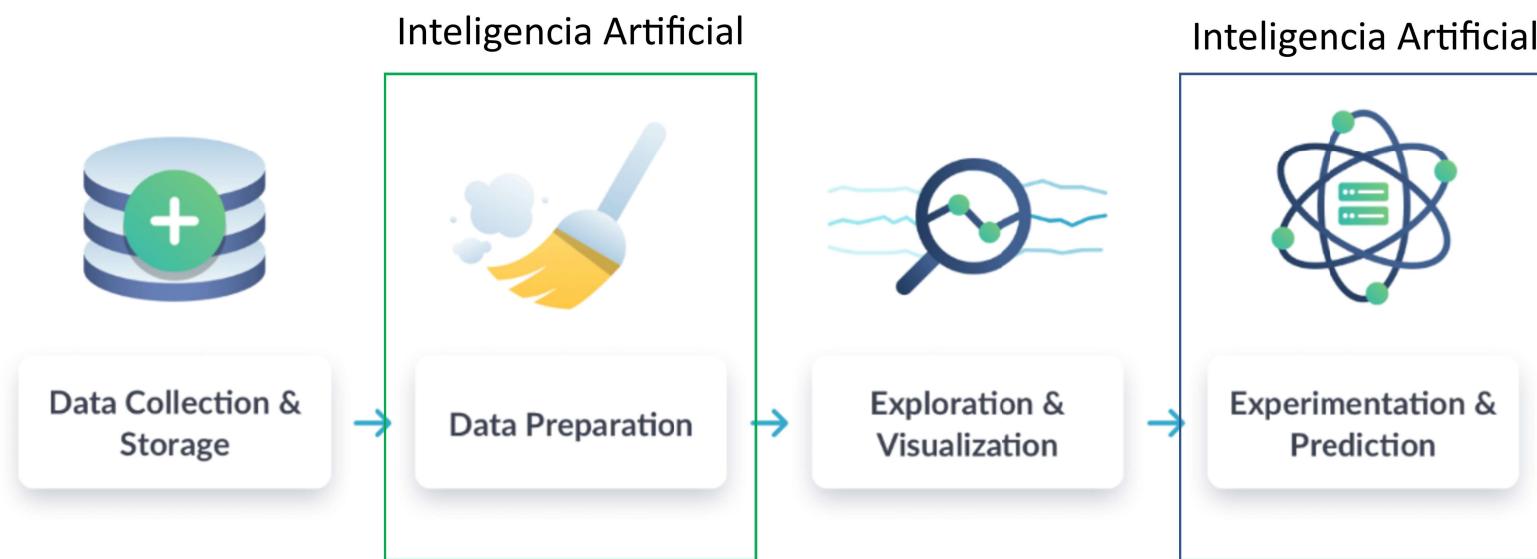
- **Big Data** (o macrodatos) se refiere a enormes volúmenes de datos que no pueden procesarse de manera efectiva con las aplicaciones tradicionales.
- El término se suele relacionar con **ciencia de datos**, pues esa suele ser su fuente de información para análisis; que puede realizarse con IA o otras metodologías.
- La ciencia de datos se centra en analizar los grandes **conjuntos de datos desordenados e incompletos**, para llegar a hallazgos que impulsan decisiones sobre operaciones y productos.

# Big data

**El análisis de Big Data ayuda a las organizaciones a aprovechar sus datos y utilizarlos para identificar nuevas oportunidades.** Eso, a su vez, conduce a movimientos de negocios más inteligentes, operaciones más eficientes, mayores ganancias y clientes más felices.

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources. 	The types of data: structured, semi-structured, unstructured. 	The speed at which big data is generated. 	The degree to which big data can be trusted. 	The business value of the data collected. 	The ways in which the big data can be used and formatted. 

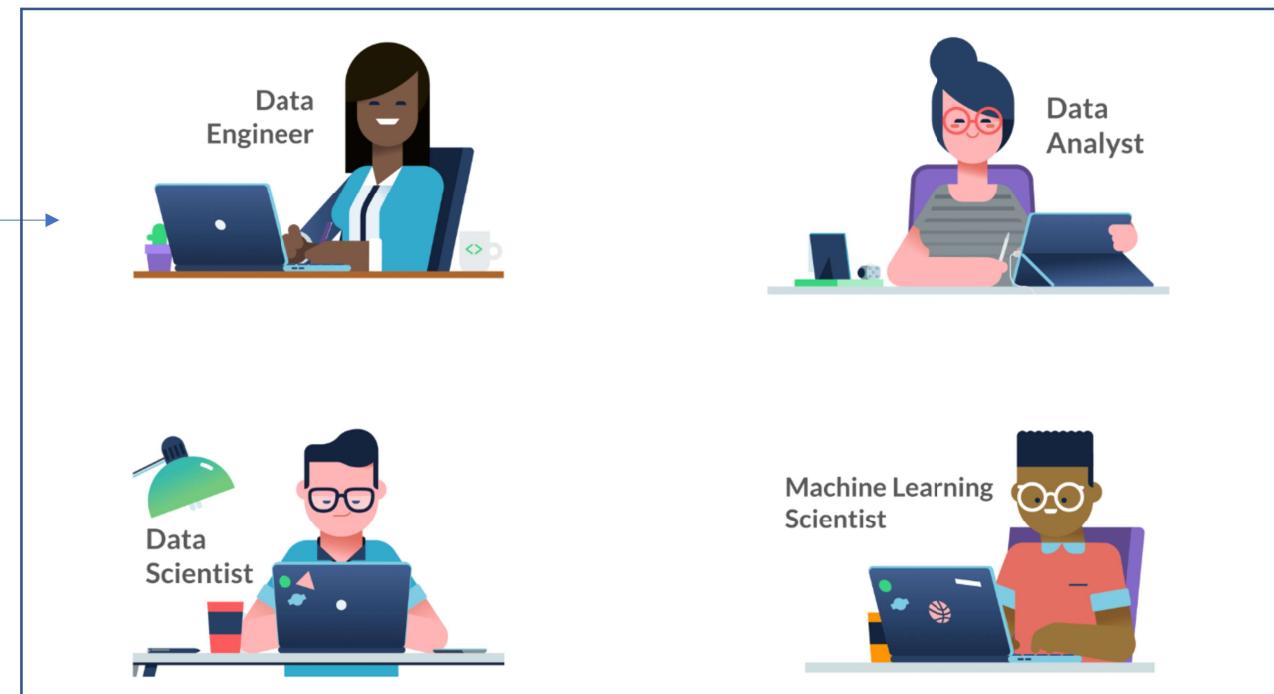
# Flujo de trabajo de la ciencia de datos



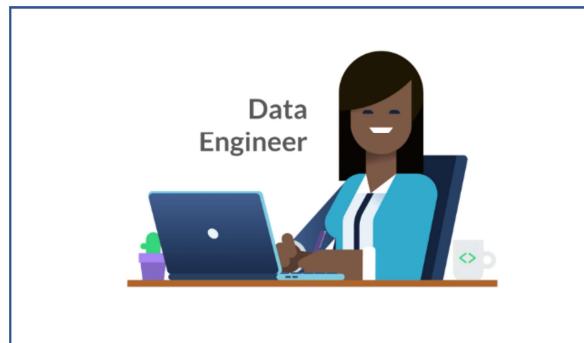
# Roles en la ciencia de datos



Proponer proyectos para aplicar IA en una organización



# Roles en la ciencia de datos

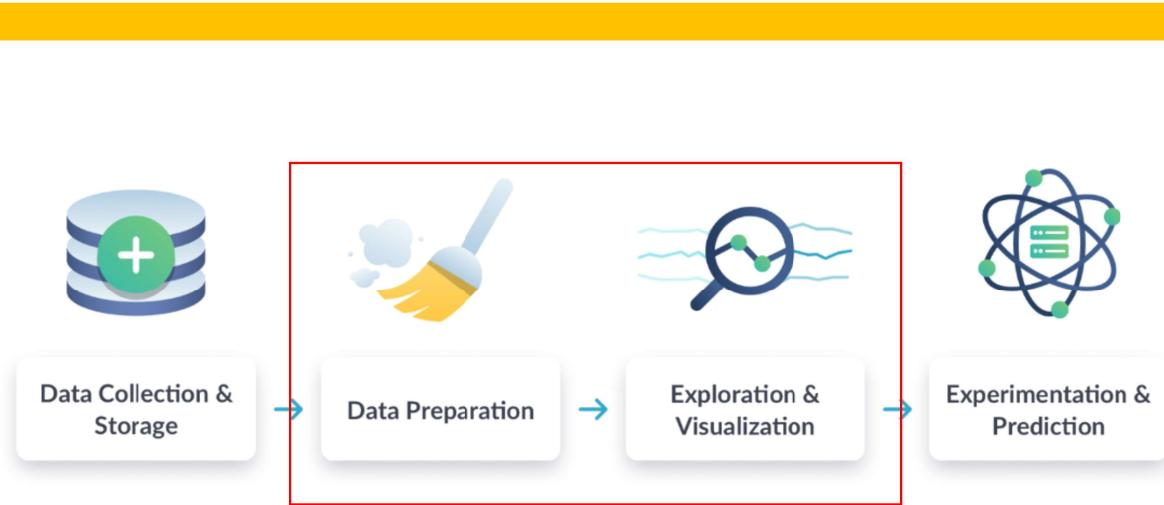


Se encargan principalmente de crear y mantener soluciones para obtener y almacenar los datos



**Conocimientos necesarios:**  
**Clave:** SQL y otros tipos de bases de datos  
**Opcional:** Procesado de datos (Python, Scala, Java)  
**Salud:** Conocimiento de sistemas HIS, HCE, PACS,...

# Roles en la ciencia de datos



## Preparación de datos:

Crean informes y dashboards para dar información preliminar sobre los datos.

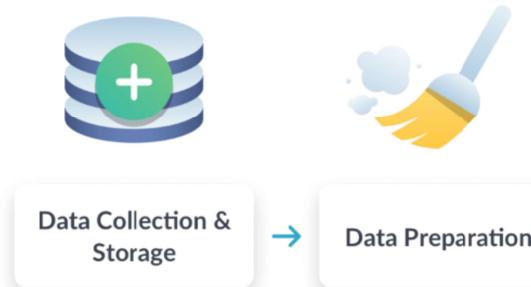
Análisis sencillos

## Conocimientos necesarios:

**Clave:** Obtener datos de bases de datos SQL, Excel, Tableau, Power BI

**Opcional:** Otras herramientas de análisis

# Roles en la ciencia de datos



Análisis exploratorio: estadística, análisis de datos, métodos básicos de IA (machine learning)

**Conocimientos necesarios:**

**Clave:** Python (Scipy, scikit-learn)

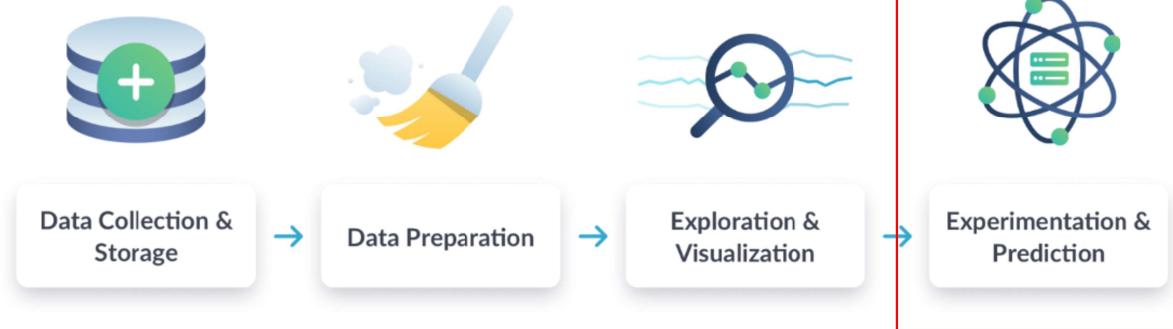
**Opcional:** Obtener datos de bases de datos SQL

**Salud:** NLP, Extracción de biomarcadores de imagen, etc...

# Roles en la ciencia de datos



Predicción de eventos futuros y programación de algoritmos para aplicaciones.



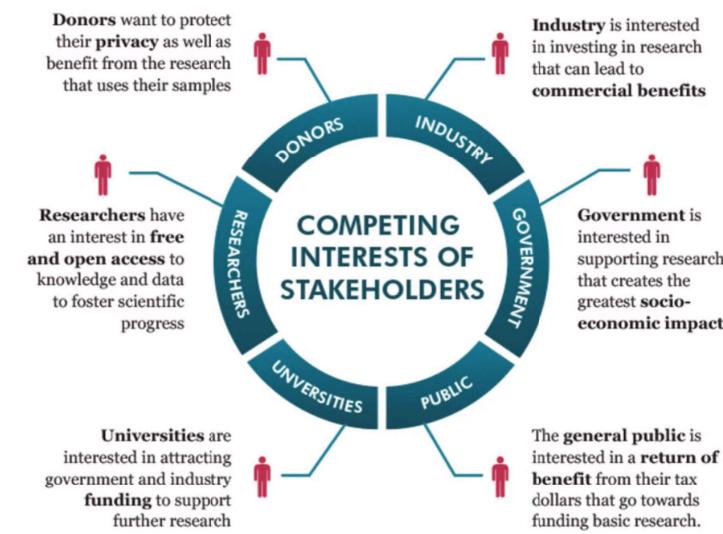
**Conocimientos necesarios:**  
**Clave:** Pytorch, Tensorflow, Keras



# Ciencia de datos en Sanidad

# Ciencia de datos en Sanidad

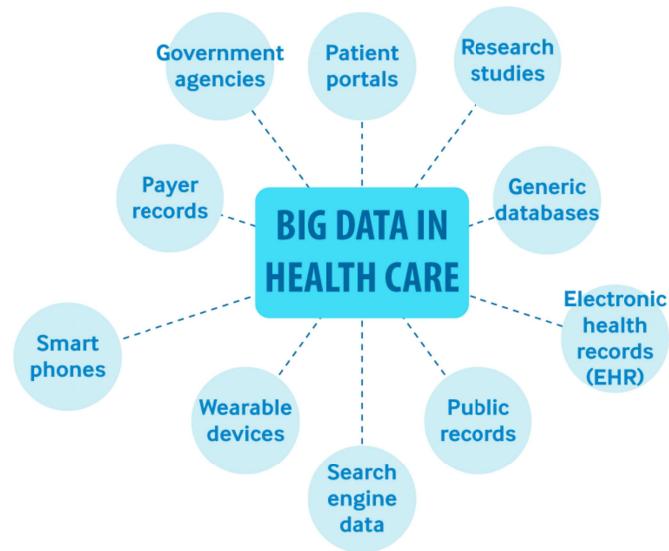
- El ecosistema sanitario se está dando cuenta de la importancia del Big Data y de las aplicaciones de la Inteligencia artificial, **que puede aportar mejoras a prácticamente cualquier proceso dentro de la operativa sanitaria de rutina.** La generalización de los datos aportaría grandes beneficios a todos los actores en la cadena de valor.



# ¿Por qué ahora?

En el mundo moderno, las computadoras y los teléfonos inteligentes se han apoderado de muchos aspectos de nuestras vidas, y la industria de la salud no es una excepción. Los profesionales de la salud están migrando cada vez más los datos de salud y atención médica del papel a los formatos electrónicos, y las instalaciones de atención médica están generando cantidades masivas de datos como resultado.

Sources of Big Data in Health Care



# Limitaciones

- Sin embargo, los casos de éxito han sido hasta hoy limitados, en investigaciones académicas o de investigación.
- Las tendencias en la industria de la salud hacia una mayor recolección de datos, la computación en la nube y el aprendizaje automático crean un ambiente propicio para que las soluciones de Big Data tengan un gran impacto en el futuro inmediato, pero existen limitaciones significativas.

# Limitaciones

- **Datos incompletos o de baja calidad:** La calidad de los resultados de la IA depende en gran medida de la calidad de los datos con los que se entrena. La falta de datos completos y precisos puede limitar la efectividad de la IA en la salud.



# Limitaciones

- **Datos heterogéneos y no estructurados:** La mayoría de los datos médicos son heterogéneos y no estructurados, lo que dificulta su procesamiento y análisis por parte de la IA.

## Datos no estructurados

- El **Paciente 1**, Juan Pérez tiene 18 años. Su temperatura corporal era de 36.2.
- El **Paciente 2**, Antonio Rodríguez, tiene 36 años, y mostró una lectura de temperatura de 37.5.
- Por último, el **Paciente 3**, Joaquín López, tiene 43 años y fue el único que mostró fiebre alta, con una lectura de 38.2.

## Datos semi-estructurados

```
<Hospital 1>
<Paciente ID =1>
<Nombre> Juan Pérez </Nombre>
<Edad> 18 </ Edad >
<Temp> 36.2 </ Edad >
<Paciente ID =2>
<Nombre> Antonio Rodríguez </Nombre>
<Edad> 36 </ Edad >
<Temp> 37.5 </ Edad >
<Paciente ID =3>
<Nombre> Joaquín López </Nombre>
<Edad> 43 </ Edad >
<Temp> 38.2 </ Edad >
```

## Datos estructurados

ID	NOMBRE	EDAD	TEMPERATURA
1	Juan Pérez	18	36.2
2	Antonio Rodríguez	36	37.5
3	Joaquín López	43	38.2

# Limitaciones

---

- **Falta de interoperabilidad de los sistemas de información:**  
Muchos sistemas médicos utilizan formatos y estándares diferentes para almacenar y compartir datos, lo que dificulta la integración y la recopilación de datos para su uso en la IA.



# Limitaciones

---

- **Protección de la privacidad de los datos:** La privacidad y la seguridad de los datos médicos son preocupaciones críticas en la aplicación de la IA en la salud, y es importante garantizar que los datos se manejen de manera responsable y conforme a las regulaciones aplicables.



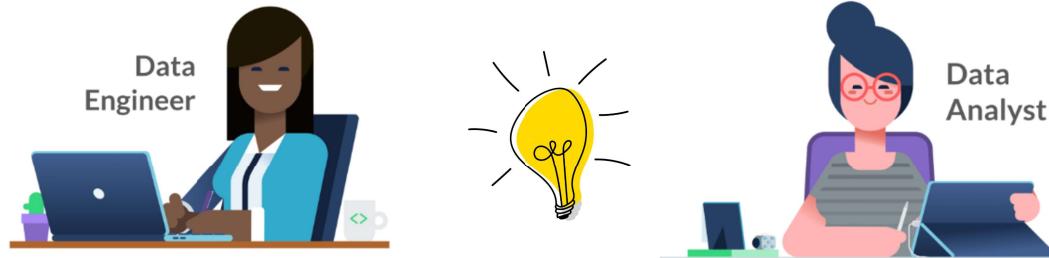
# Big data

## Volviendo a las 6 Vs

VOLUME	VARIETY	VELOCITY	VERACITY	VALUE	VARIABILITY
The amount of data from myriad sources. 	The types of data: structured, semi-structured, unstructured. 	The speed at which big data is generated. 	The degree to which big data can be trusted. 	The business value of the data collected. 	The ways in which the big data can be used and formatted. 

# La solución

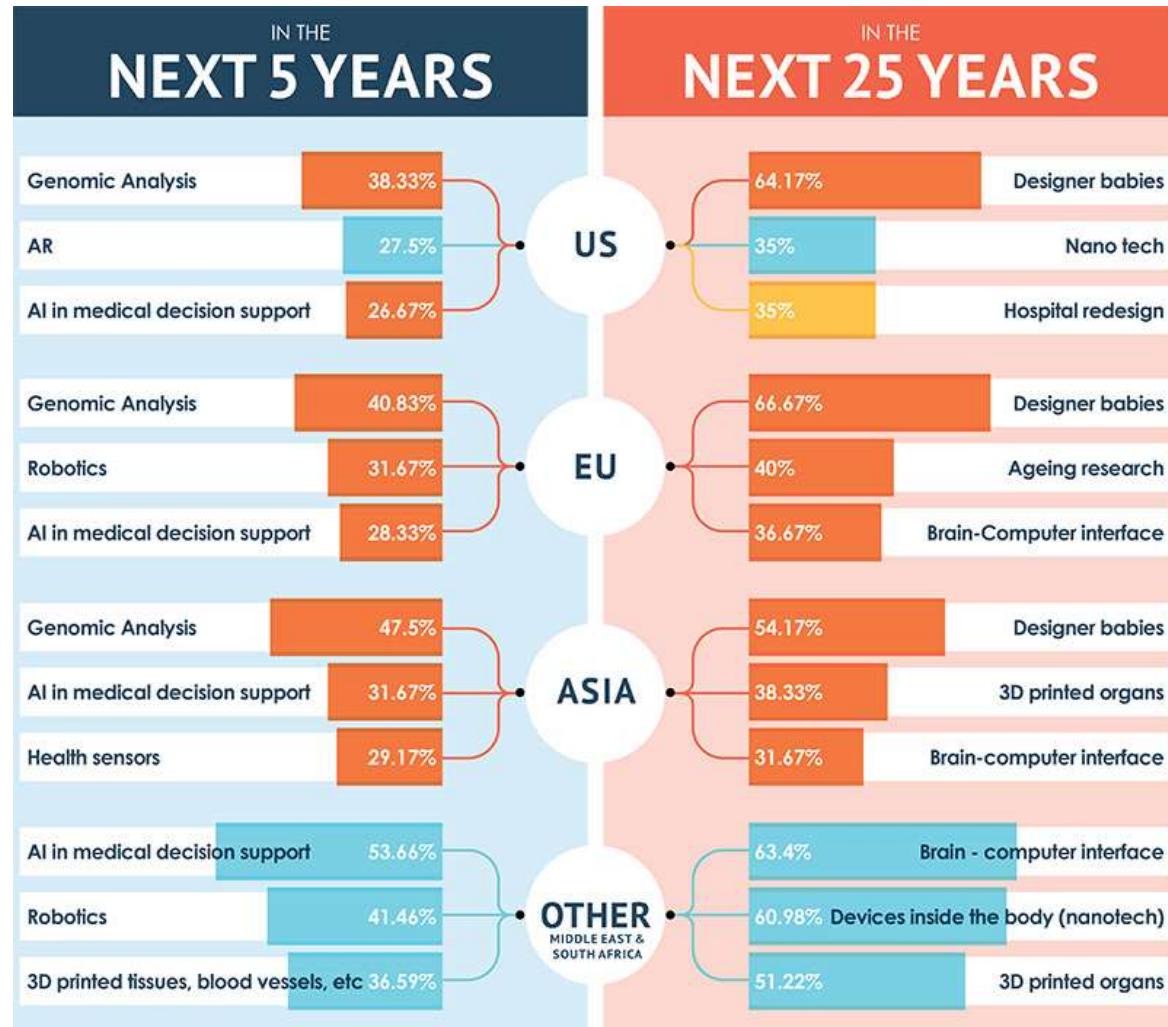
- Muchos hospitales y sistemas de salud están incorporando **figuras de gestión relacionadas con la ciencia de datos**. La principal misión de estas figuras es dar valor a los datos generados a través de la implementación de soluciones de IA.
- Se ha dado un impulso a los estándares de interoperabilidad para facilitar el uso de datos
- Bases de datos sanitarias.



# El futuro

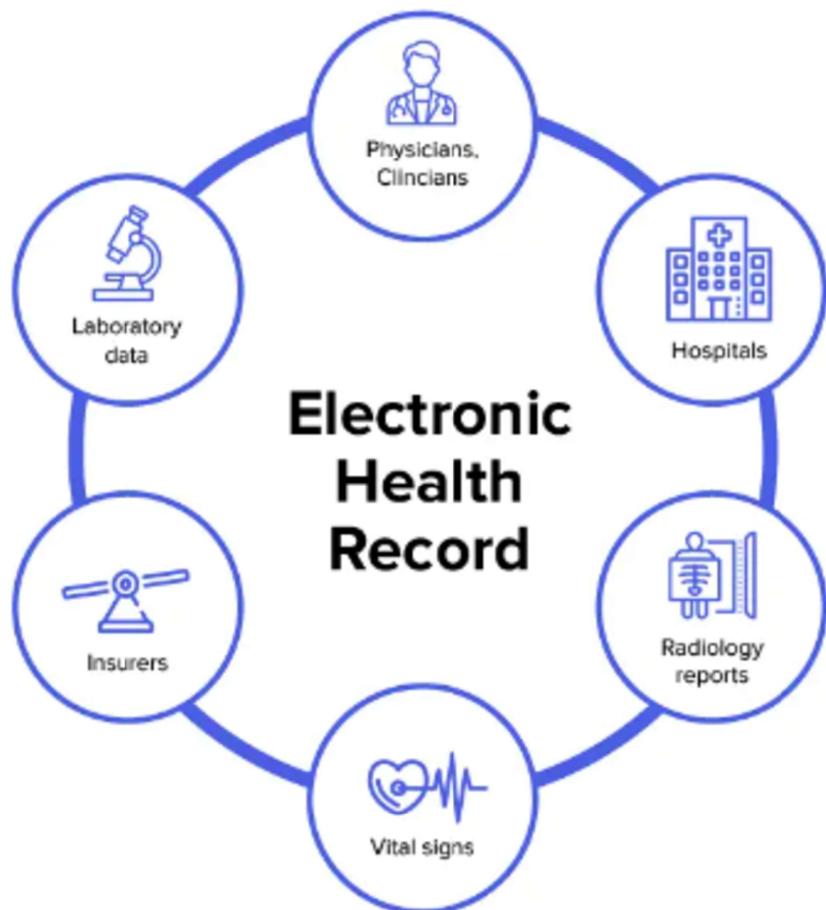
• El COVID-19 ha impulsado un ritmo de digitalización sin precedentes, y la atención y el monitoreo remotos han demostrado ser pilares críticos de los sistemas de salud en todo el mundo.

En los próximos años habrá un cambio fundamental de "atención médica" a "salud". Y aunque la enfermedad nunca se eliminará por completo, a través de la ciencia, los datos y la tecnología, podremos identificarla antes, intervenir de manera proactiva y comprender mejor su progresión para ayudar a los consumidores a mantener su bienestar de manera más efectiva y activa.





# Fuentes de datos



# Historia Clínica Electrónica

- Los EHR almacenan información como datos demográficos del paciente, datos clínicos, resultados de diagnóstico y planes de tratamiento. Brindan una descripción general completa del historial de salud de un paciente, lo que permite a los proveedores de atención médica tomar decisiones informadas sobre el tratamiento.
- Contienen tanto datos estructurados como no estructurados
- Los datos estructurados en los EHR generalmente se ingresan en un formato específico, como menús desplegables o casillas de verificación. Los datos no estructurados, como notas médicas, pueden ingresarse como texto libre. El desafío de utilizar datos no estructurados es extraer información significativa e integrarla con los datos estructurados para brindar una visión integral del historial médico de un paciente.

# Wearables



Los wearables se han vuelto cada vez más populares en los últimos años y tienen el potencial de revolucionar la atención médica al proporcionar datos en tiempo real sobre el estado de salud de un paciente.

Variedad de métricas de salud: frecuencia cardíaca, actividad física, patrones de sueño o niveles de estrés. Estos datos se pueden integrar con registros de salud electrónicos (EHR) para proporcionar una imagen más completa de la salud de un paciente. Además, los dispositivos portátiles se pueden usar para monitorear pacientes con afecciones crónicas, como diabetes o enfermedades cardíacas, y alertar a los proveedores de atención médica sobre posibles problemas de salud.

# Imágenes médicas



Las imágenes médicas, como las radiografías y las tomografías computarizadas, contienen información valiosa para diagnosticar y tratar a los pacientes, pero a menudo son datos no estructurados.

Para superar esto, las imágenes médicas se pueden transformar en datos estructurados extrayendo información relevante y organizándola para facilitar el análisis. Los algoritmos de aprendizaje automático se pueden usar para identificar y categorizar las características de la imagen.

# Laboratorio

 University Health Network	LABORATORY MEDICINE PROGRAM																																																		
DEPARTMENT OF PATHOLOGY 200 Elizabeth Street Toronto, Ontario, M5G 2C4 TEL: 416-340-3325 FAX: 416-586-9901																																																			
<b>Surgical Pathology Consultation Report</b> * Added *  <table border="1"><tr><td>Patient Name:</td><td>Patient, USCAP</td><td>Service:</td><td>TGH Thoracic</td><td>Accession #:</td><td>S16-12345</td></tr><tr><td>MRN:</td><td>9876543</td><td>Visit #:</td><td>23412312345</td><td>Collected:</td><td>May-05-2016</td></tr><tr><td>DOB:</td><td>11/22/1947 (Age: 68)</td><td>Location:</td><td>2C Pre Operative Care Unit</td><td>Received:</td><td>May-05-2016</td></tr><tr><td>Gender:</td><td>F</td><td>Facility:</td><td>TGH/PMH</td><td>Reported:</td><td>Jun-01-2016</td></tr><tr><td>HCN:</td><td>123456775CH</td><td></td><td></td><td></td><td></td></tr><tr><td>Ordering MD:</td><td>Deep Cutter, MD</td><td></td><td></td><td></td><td></td></tr><tr><td>Copy To:</td><td>Good P Friend, MD</td><td></td><td></td><td></td><td></td></tr><tr><td></td><td>Stat Response, MD</td><td></td><td></td><td></td><td></td></tr></table>				Patient Name:	Patient, USCAP	Service:	TGH Thoracic	Accession #:	S16-12345	MRN:	9876543	Visit #:	23412312345	Collected:	May-05-2016	DOB:	11/22/1947 (Age: 68)	Location:	2C Pre Operative Care Unit	Received:	May-05-2016	Gender:	F	Facility:	TGH/PMH	Reported:	Jun-01-2016	HCN:	123456775CH					Ordering MD:	Deep Cutter, MD					Copy To:	Good P Friend, MD						Stat Response, MD				
Patient Name:	Patient, USCAP	Service:	TGH Thoracic	Accession #:	S16-12345																																														
MRN:	9876543	Visit #:	23412312345	Collected:	May-05-2016																																														
DOB:	11/22/1947 (Age: 68)	Location:	2C Pre Operative Care Unit	Received:	May-05-2016																																														
Gender:	F	Facility:	TGH/PMH	Reported:	Jun-01-2016																																														
HCN:	123456775CH																																																		
Ordering MD:	Deep Cutter, MD																																																		
Copy To:	Good P Friend, MD																																																		
	Stat Response, MD																																																		
<b>Specimen(s) Received</b> 1. Lymph-Node: ST10R TB Angle 2. Right middle lobe 3. Station 11R 4. Station 4R 5. Station 7 6. Interlobar ST11 7. Right middle and upper bilobectomy																																																			
<b>Consolidated Theranostic Report</b>																																																			
<b>Interpretation</b>																																																			
Invasive moderately differentiated adenocarcinoma, acinar-predominant, pT2aN1 - POSITIVE for EGFR L858R mutation (see Molecular Diagnostics report) - NEGATIVE for ALK by immunohistochemistry (performed using the 5A4 antibody with a protocol optimized for detection of ALK gene rearrangement) - See Diagnosis, Comment, and Synoptic Report below for further details																																																			
Signed out by: Lung Path, MD Date Reported: Jun-01-2016																																																			
<b>Diagnosis</b>																																																			
1,3-6. Lymph nodes (ST10R right tracheobronchial, ST11R right interlobar, ST4R right lower paratracheal, ST7 subcarinal, ST11 interlobar): - At least one lymph node per station, negative for malignancy (x5) (0/5)																																																			
2. Lung, resection (right middle lobectomy): a. Invasive moderately differentiated adenocarcinoma, acinar-predominant, pT2aN1, with: i. Greatest tumor dimension: 1.2 cm (see Comment) ii. Visceral pleural and lympho-vascular invasion present iii. Stapled parenchymal resection margin positive for carcinoma (see Comment) b. One of five lymph nodes focally positive for adenocarcinoma by direct invasion (1/5) (see Comment)																																																			
Patient, USCAP		Page 1 of 5																																																	

Los resultados de laboratorio juegan un papel crucial en el diagnóstico y seguimiento de la salud del paciente. Sin embargo, los resultados de laboratorio a menudo se presentan en formatos no estructurados, lo que dificulta su análisis.

Para superar este desafío, los resultados de laboratorio se pueden transformar en datos estructurados al estandarizar el formato y organizar la información de una manera que permita un fácil análisis y comparación con otros datos de pacientes. Esto se puede lograr mediante el uso de técnicas de normalización y extracción de datos automatizadas, como el procesamiento del lenguaje natural (NLP).



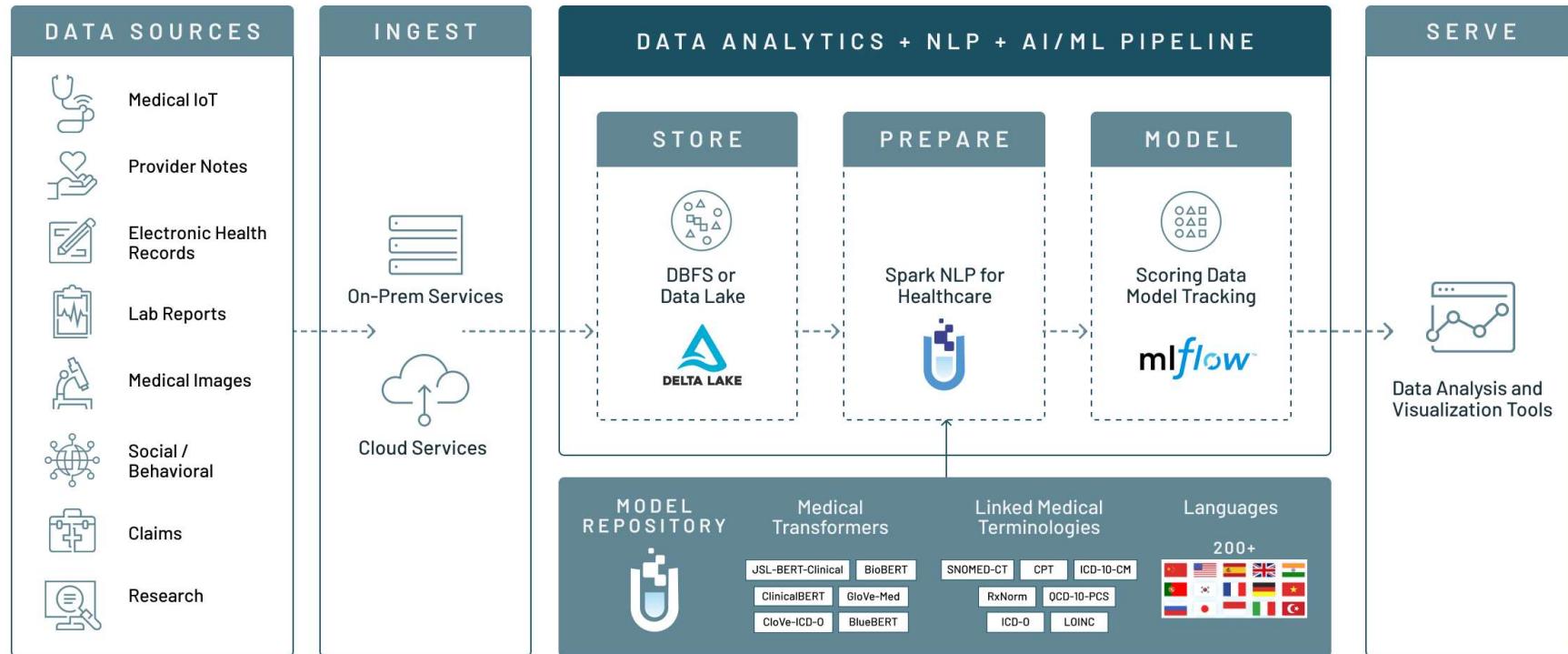
# Procesado de datos

# Natural Language Processing

- Los registros médicos, los comentarios de los pacientes, las evaluaciones del desempeño de los médicos o incluso los comentarios de las redes sociales, pueden ser una rica fuente de datos para ayudar en la toma de decisiones clínicas y la mejora de la calidad.
- Apróximadamente el 80% de la información en los servidores hospitalarios se encuentra en un formato de texto libre. A menudo, estos conjuntos de datos de texto abierto son tan grandes que no sería práctico sintetizar manualmente toda la información útil con técnicas de investigación cualitativas.
- El procesamiento del lenguaje natural (NLP) describe un conjunto de técnicas utilizadas para convertir pasajes de texto escrito en conjuntos de datos interpretables que pueden analizarse mediante modelos estadísticos y de aprendizaje automático.

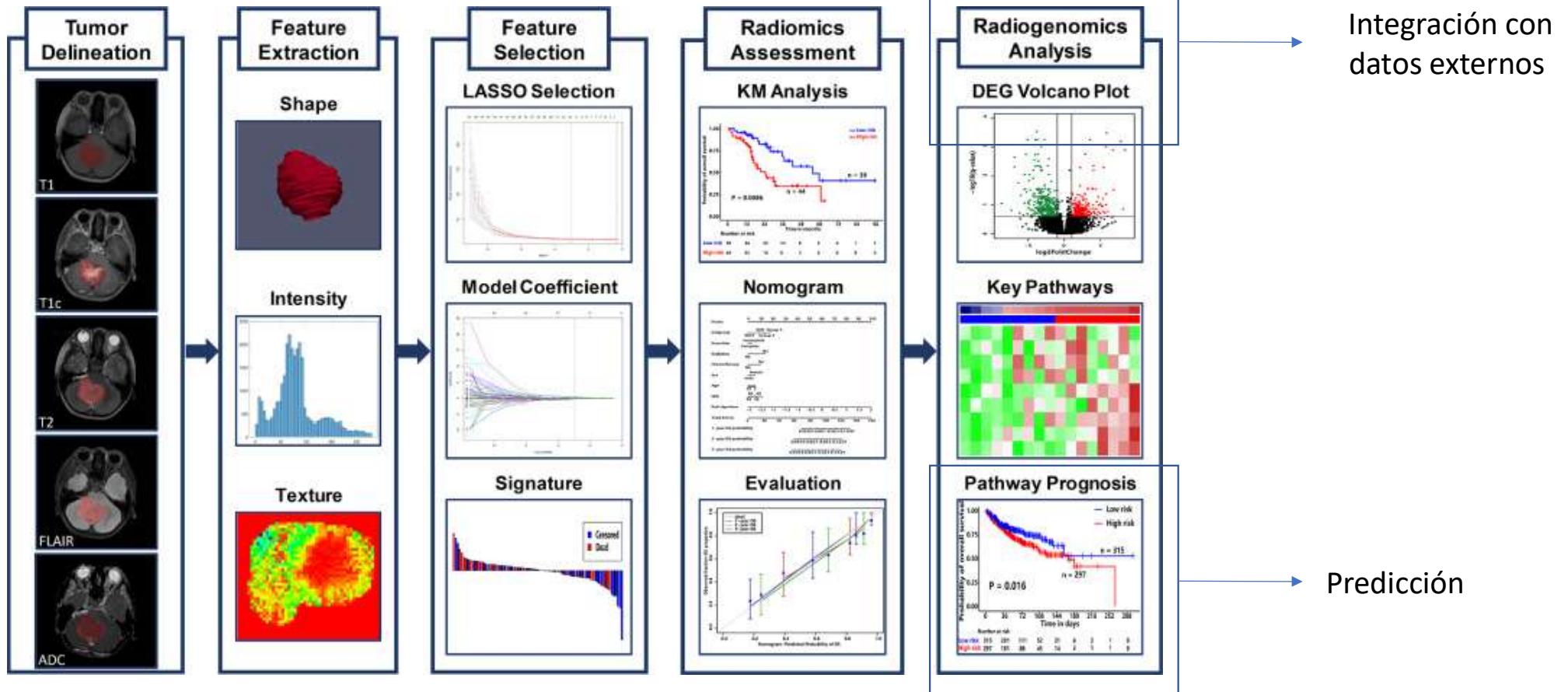


# Natural Language Processing

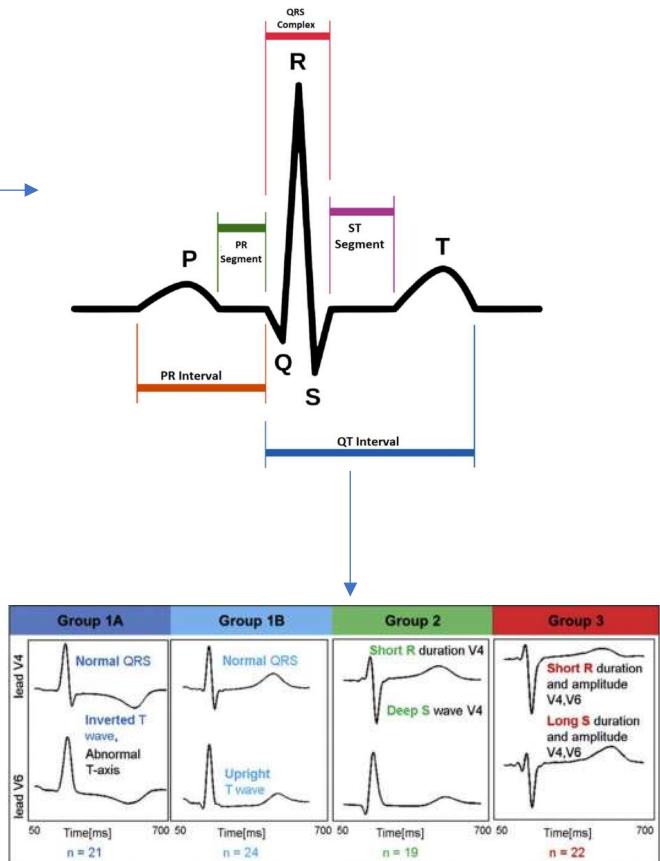
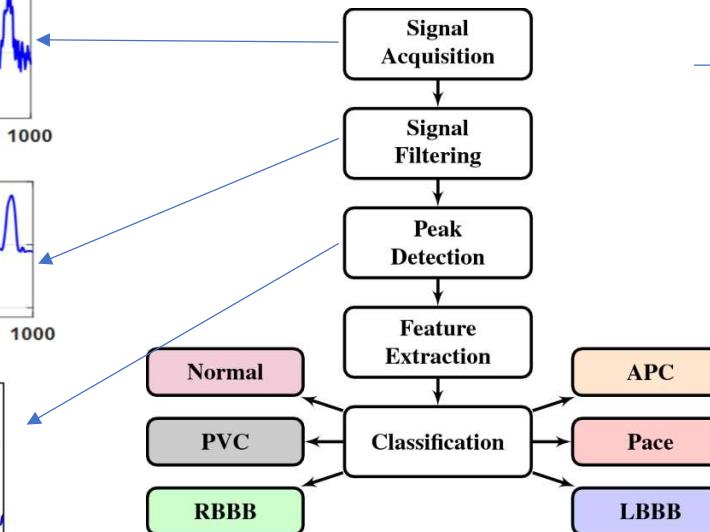
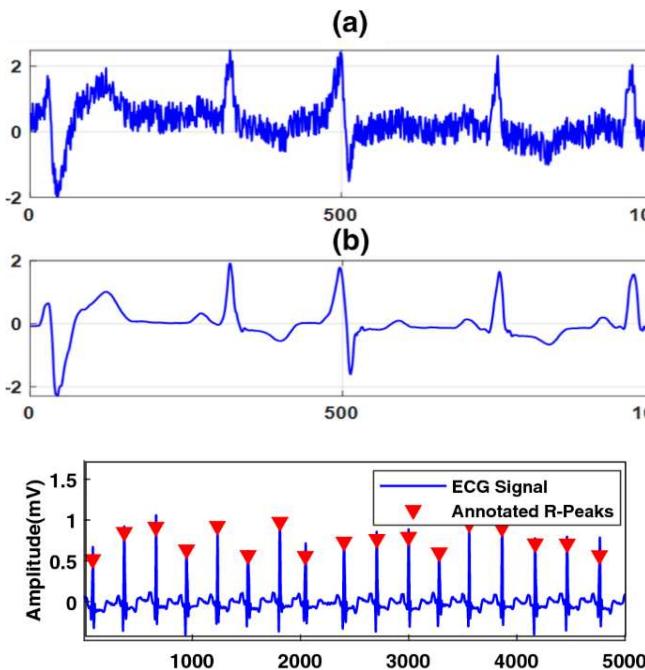


[Applying Natural Language Processing to Healthcare Text at Scale - The Databricks Blog](#)

# Procesado de imagen

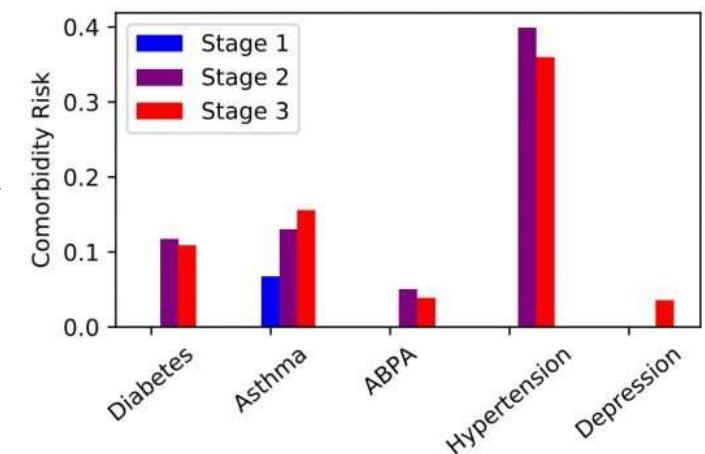
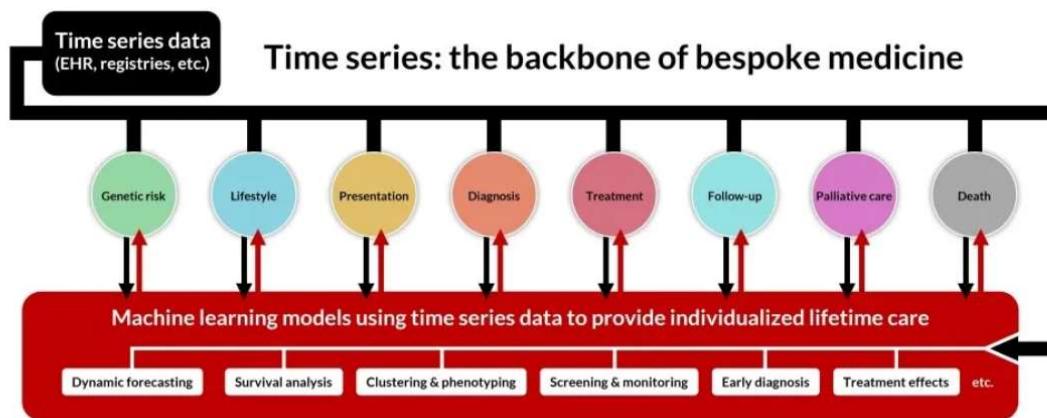


# Datos con dimensión temporal



# Series de datos

Los conjuntos de datos de series temporales, como los registros electrónicos de salud (EHR), representan fuentes de información valiosas (pero imperfectas) que abarcan toda la vida de atención de un paciente. Pueden capturar los riesgos genéticos y de estilo de vida, señalan la aparición de enfermedades, muestran el advenimiento de nuevas morbilidades y comorbilidades, indican el momento y la etapa del diagnóstico y documentan el desarrollo de planes de tratamiento, así como su eficacia.



Data  
Saves  
Lives



GRACIAS POR VUESTRA  
ATENCIÓN