

10 Ciencia de Datos y Tecnologías Emergentes

Índice de Contenidos del Módulo

10	Ciencia de Datos y Tecnologías Emergentes	1
10 - 1	Objetivos humanitarios a través de la Ciencia de Datos	7
10 - 2	Rol de un equipo de voluntarios en materia de Ciencia de Datos	11
10 - 3	Cómo sería la descripción del puesto de trabajo de un Científico de Datos	12
10 - 4	Procesos de implementación sostenibles y mejora de la efectividad	13
10 - 5	Datos, Tecnologías Emergentes y Objetivos Sostenibles	15
10 - 6	Diversidad Social, Equidad e Inclusión	21
10 - 7	Mapeo de Aspectos del Trabajo en materia de Ciencia de Datos	26
10 - 8	¿Por qué necesitamos la Ciencia de Datos?	29
10 - 9	Decisiones de ayuda en contextos específicos	33
10 - 10	Visión general de la Plataforma y los Procesos típicos de la Ciencia de Datos	38
10 - 11	Cómo implementar un modelo de Ciencia de Datos	49
10 - 12	Análisis de Churn de la Cruz Roja: Jupyter Notebook	52

Ciencia de Datos y Tecnologías Emergentes

Este módulo introduce a los equipos en los conceptos básicos de la ciencia de datos y su conexión con los objetivos y metas humanitarias. Su objetivo es guiar a los equipos que deseen explorar las posibilidades de la ciencia de datos y las tecnologías emergentes en su trabajo diario hacia la innovación y la transformación.

Cuestiones que explora este módulo

- ▶ ¿Cómo podemos adoptar y utilizar eficazmente la ciencia de datos y las nuevas tecnologías en nuestro trabajo?
- ▶ ¿Cómo podemos entender y apoyar el uso de la ciencia de datos y las tecnologías emergentes en nuestro trabajo?

Objetivos de Aprendizaje

- ▶ Explorar cómo los equipos entienden y utilizan la ciencia de datos y las tecnologías emergentes en su trabajo.
- ▶ Guiar a los equipos hacia los factores críticos de éxito que deben conocer y las preguntas clave que deben plantearse al emprender proyectos de ciencia de datos.
- ▶ Evaluar el nivel necesario de sensibilización sobre las tecnologías emergentes

Temas del Módulo

- ▶ Por qué utilizar tecnologías emergentes en nuestro trabajo
- ▶ Cómo utilizar los datos y los principios fundamentales para abordar los retos y las transformaciones humanitarias estratégicas
- ▶ ¿Cuáles son las iniciativas y perspectivas de la ciencia de datos y las tecnologías emergentes en nuestro trabajo?

Recetas

Proceso paso a paso sugerido para alcanzar los objetivos de aprendizaje

- 1 Comience por identificar los **Objetivos humanitarios a través de la Ciencia de Datos (10 - 1)** para comprender por qué se introducen las tecnologías emergentes en la labor humanitaria.
- 2 A continuación, debata los **Procesos de implementación sostenibles y mejora de la efectividad (10 - 4)** que se muestra en el lienzo (canvas) de aprendizaje de tecnologías emergentes.
- 3 Examine detenidamente **Datos, Tecnologías Emergentes y Objetivos Sostenibles (10 - 5)** y cómo se representa en el marco de los objetivos sostenibles.
- 4 Permita a los grupos hacer **Mapeo de Aspectos del Trabajo en materia de Ciencia de Datos (10 - 7)** que encuentran en su trabajo y discutir por qué necesitan aprendizaje automático.
- 5 Oriente a los participantes sobre **Cómo implementar un modelo de Ciencia de Datos (10 - 11)**.
- 6 Presente a los grupos con **Decisiones de ayuda en contextos específicos (10 - 9)**.

- 7 En un debate grupal, pida a los participantes que compartan si incorporarán proyectos de ciencia de datos para ayudar en su labor humanitaria.

Ingredientes

Elija los ingredientes para crear su propia receta. ¿Tiene algún ingrediente que nos falte? Envíe un correo electrónico a data.literacy@ifrc.org

Ejercicios

Presentaciones a utilizar y/o adaptar:

- ▶ Oportunidades para abordar los objetivos humanitarios a través de la ciencia de datos
- ▶ Marco de referencia para los objetivos de sostenibilidad: una mirada más atenta a la sostenibilidad en la IA
- ▶ Principios: diversidad social, equidad, inclusión en la ciencia de datos
- ▶ Mapeo de Aspectos del Trabajo en materia de Ciencia de Datos
- ▶ Por qué necesitamos aprendizaje automático
- ▶ Preguntas para orientar las decisiones sobre plataformas, herramientas y selección de competencias
- ▶ Cómo implementar un modelo de Ciencia de Datos

Presentaciones con diapositivas

Borrador de presentaciones para revisar/utilizar:

- ▶ Procesos de implantación sostenibles para mejorar la madurez. Proporciona contexto sobre el lienzo (canvas) de aprendizaje de las tecnologías emergentes.

Listas de Verificación/Folletos/Materiales

Para documentar los elementos esenciales de la experiencia de aprendizaje:

- ▶ Función de un equipo voluntario de ciencia de datos (Canvas). Material para utilizar con **Por qué necesitamos la Ciencia de Datos (10 - 8)** (ejercicio).
- ▶ Cómo debe ser la descripción de puesto de un científico de datos (Términos de Referencia). Material para utilizar con **Por qué necesitamos la Ciencia de Datos (10 - 8)** (ejercicio).

- ▶ Plataformas de ciencia de datos (Diapositivas). Proporciona contexto sobre los componentes granulares que se incluyen en las herramientas de ciencia de datos y su uso. Se puede hacer referencia a este material en **Mapeo de Aspectos del Trabajo en materia de Ciencia de Datos (10 - 7)** (ejercicio).
- ▶ Red Cross Churn Analysis (Jupyter Notebook). Material para utilizar con **Cómo implementar un modelo de Ciencia de Datos (10 - 11)** (ejercicio).

Recursos Externos

- ▶ [An Abstraction Framework for Reducing Complexity in AI Governance \(Article\)](#)
- ▶ [Knowledge sharing and collaboration: Example of data and digital week](#) (Post en Blog)
- ▶ [Examples from Red Cross Red Crescent experiences: Go Github data science projects](#) (Repositorio GitHub)
- ▶ [Artificial Intelligence Suitability Framework](#) (Diapositivas)
- ▶ Alsolbi I., et.al (2021). Analyzing Donors Behaviors in Nonprofit Organisations: A Design Science Research Framework.
- ▶ Goffi E. R., Momcilovic A. (dir.) (2022). [Global Trends in AI 2022: Food for thought from GAIEI experts. Global AI Ethics Institute](#), Notes nº 1.

Próximos Pasos

Módulos relacionados en el Data Playbook con contenido sugerido

Consulte el Módulo **Prácticas Responsables y Protección de Datos (7)** para comprender mejor cómo mitigar los riesgos y los **Atajos mentales y errores de razonamiento (6 - 3)** (Folleto) en el Módulo **Comprender y Analizar Datos (6)**.

Crédito

Mahendra Samarawickrama, Jenny Paola Yela-Bello, y colaboradores IFRC V1 Sprint and Data Playbook

10 - 1 Objetivos humanitarios a través de la Ciencia de Datos

Al aplicar los datos y las tecnologías emergentes relacionadas a la humanidad, es importante que hayamos identificado las oportunidades, la ética, los valores y la justicia social. Desde esta perspectiva, los equipos deben asegurarse de que entienden y comprenden por qué la ciencia de datos y la tecnología emergente se utilizan en su trabajo. Este ejercicio permitirá a los equipos examinar y debatir de manera conjunta algunos marcos de referencia. Estos marcos de referencia ayudan a los equipos a identificar oportunidades que abordan objetivos humanitarios específicos dentro de su trabajo.

- ▶ **Personas:** 2 a 8 personas
- ▶ **Tiempo:** 40 Minutos
- ▶ **Dificultad:** Media
- ▶ **Materiales Virtual:** plataforma de reunión virtual, documento/ espacio de escritura compartido
- ▶ **Materiales Presencial:** Rotafolios/pizarras, notas adhesivas, rotuladores

Ejercicio

Introducción

En este ejercicio abordamos por qué es importante tomar en consideración los datos, la IA y las tecnologías emergentes desde la perspectiva humanitaria. El objetivo del ejercicio es encontrar respuestas a por qué los datos, la IA y las tecnologías emergentes deben considerarse un componente de la información estratégica, y por qué desempeñan un papel importante en la labor humanitaria.

Parte 1: Explorar

En equipos, por favor revisen lo siguiente



Para este ejercicio, se utilizan como marcos de referencia la estrategia 2030 de la FICR y los 17 Objetivos de Desarrollo Sostenible de las Naciones Unidas. Los equipos pueden trabajar

grupos para debatir cuestiones sobre por qué la ciencia de datos y la tecnología emergente deben ser un componente clave en la formación sobre estrategias. Tras los debates, el facilitador puede integrar las ideas compartidas en una visión más amplia sobre por qué las oportunidades de la ciencia de datos son aplicables a su trabajo y cómo pueden aplicarse.

Nota: [The role of artificial intelligence in achieving the Sustainable Development Goals.](#)

Parte 2: Debatir

Pregunta 1: Teniendo en cuenta la Estrategia 2030, identifique las oportunidades que la ciencia de datos y la Inteligencia Artificial aportan a la humanidad.

Puede debatirse durante 10 minutos

[Estrategia 2030 de la FICR](#)

3 Objetivos

1
Personas capaces de anticipar las crisis, superarlas y recuperarse rápidamente de ellas

2
Personas que gocen de condiciones de vida sana y segura, dignidad y oportunidades de prosperar

3
Personas que se movilicen en pro de comunidades inclusivas y pacíficas



5 Desafíos

1
Crisis relacionadas con el clima y el medio ambiente

2
Crisis y desastres evolutivos

3
Desfases crecientes en materia de salud y bienestar

4
Migración e identidad

5
Valores, poder e inclusión

7 Transformaciones

1
Apoyo a las Sociedades Nacionales y desarrollo de estas en calidad de agentes locales sólidos y eficientes

2
Inspiración para prestar servicio voluntario y movilización de voluntarios

3
Garantía de confianza y rendición de cuentas

4
Labor eficaz en calidad de red distribuida

5
Incidencia en la labor humanitaria

6
Transformación digital

7
Financiación con miras al futuro

Pregunta 2: ¿Conoce los **17 Objetivos de Desarrollo Sostenible (ODS) de las Naciones Unidas**? ¿Sabe que la IA puede contribuir al 79% de estos objetivos? ¿Puedes relacionar algunos de ellos con nuestra misión? A continuación, el equipo puede debatir el diagrama durante 10 minutos.



Crédito

Para el desarrollo del ejercicio Mahendra Samarawickrama

10 - 2 Rol de un equipo de voluntarios en materia de Ciencia de Datos

[Recursos para descargar](#)

10 - 3 Cómo sería la descripción del puesto de trabajo de un Científico de Datos

[Recursos para descargar](#)

10 - 4 Procesos de implementación sostenibles y mejora de la efectividad

Lienzo de aprendizaje de tecnologías emergentes: Herramienta para informar sobre el desafío empresarial

Problema y oportunidad

Propuesta de valor para el donante/visión del proyecto: Define “qué y por qué” lo hacemos.
Problema del donante: Qué problemas estamos resolviendo.
Oportunidad de negocio: Qué valor aportará a la actividad.

Segmentos de donantes o personas

Enumere sus donantes y usuarios objetivo para los que resolvemos los problemas y proporcionamos las soluciones.

Contexto de apoyo

Investigación y conocimientos existentes: Resultados clave de las pruebas iniciales con usuarios, investigación de donantes o cualquier investigación cualitativa, encuestas. También puede incluir cualquier investigación previa validada de iteraciones anteriores. Esto respalda el “qué y por qué” de la sección de *Problema y oportunidad* anterior.

Meta

Éxito: ¿Qué significa éxito y cómo sabremos que estamos aportando valor? ¿Cuál es la medida clave del éxito?

Hipótesis

Aprendizaje: La idea o solución propuesta que aportará el valor y nos acercará a la meta.
 ¿Qué estrategias clave generarán el mayor valor? ¿O qué aprendizaje validado es importante para confirmar nuestro camino hacia el éxito? Este es el “cómo” de las metas. Una buena hipótesis captura lo que necesita ser probado mediante la medición y experimentación (para ser probado o refutado).

Cronología

¿Cuáles son los hitos críticos?

Equipo/Partes interesadas

Propietario del Negocio
 Propietario del Producto

10 - 5 Datos, Tecnologías Emergentes y Objetivos Sostenibles

La misión de la Cruz Roja y de la Media Luna Roja es prevenir o reducir el sufrimiento humano, dondequiera que se encuentre.

Formamos parte del mayor movimiento humanitario del mundo, que opera en más de 192 países. Contamos con más de 14,8 millones de voluntarios en todo el mundo.

La CRMLR desempeña un rol auxiliar del gobierno - reconocido como un socio confiable para movilizar el poder de la humanidad. Con la implementación de la Estrategia 2030 y el aumento del uso de herramientas digitales en las misiones de la CRMLR, es necesario examinar más de cerca cómo las herramientas digitales (la inteligencia artificial es una de ellas) pueden ser sostenibles y reflejar los valores, fortalezas y cultura de la CRMLR.

- ▶ **Personas:** 2 a 8 personas
- ▶ **Tiempo:** 60 Minutos
- ▶ **Dificultad:** Media
- ▶ **Materiales Virtual:** plataforma de reunión virtual, documento/ espacio de escritura compartido
- ▶ **Materiales Presencial:** Rotafolios/pizarras, notas adhesivas, rotuladores

Ejercicio

En este ejercicio, el equipo, con la orientación del facilitador, hablará de los aspectos que es importante destacar y en los que hay que trabajar cuando se impulsa la IA hacia la sostenibilidad.

Parte 1: Explorar y Revisar

Será un debate en sesión plenaria dividido en 4 partes.

La sesión abarcará 4 puntos principales junto con diagramas y una visión general que puede ayudar a guiar el debate.

¿Por qué la Ciencia de Datos y la IA necesitan ser reguladas éticamente?

La Cruz Roja Australiana está trabajando en la gobernanza de la IA para la sostenibilidad porque la IA tiene un enorme potencial para la humanidad y un riesgo significativo si no se regula adecuadamente. Se prevé que la IA puede aportar hasta 15,7 billones a la economía mundial de aquí a 2030. Al mismo tiempo, la IA crea grandes riesgos para la humanidad, como las armas autónomas, la pérdida de empleos provocada por la automatización, la desigualdad socioeconómica, las violaciones de la privacidad, Deepfakes y los sesgos causados por datos y algoritmos. Además, se ha predicho que para 2022, el 85% de los proyectos de IA fracasarán debido a sesgos en los datos, los algoritmos o los equipos responsables de gestionarlos.

Una forma de abordar la IA sostenible para la sostenibilidad es mejorar la ética y la gobernanza de la IA aplicando buenas prácticas. Se le ocurren otras formas que puedan incorporarse y utilizarse?

¿Cuáles son nuestros valores, nuestro propósito y nuestra fuerza cuando nos centramos en la ciencia de datos y la IA? Abra el debate tras observar el diagrama que figura a continuación.

Vea por favor la [animación](#) whiteboard

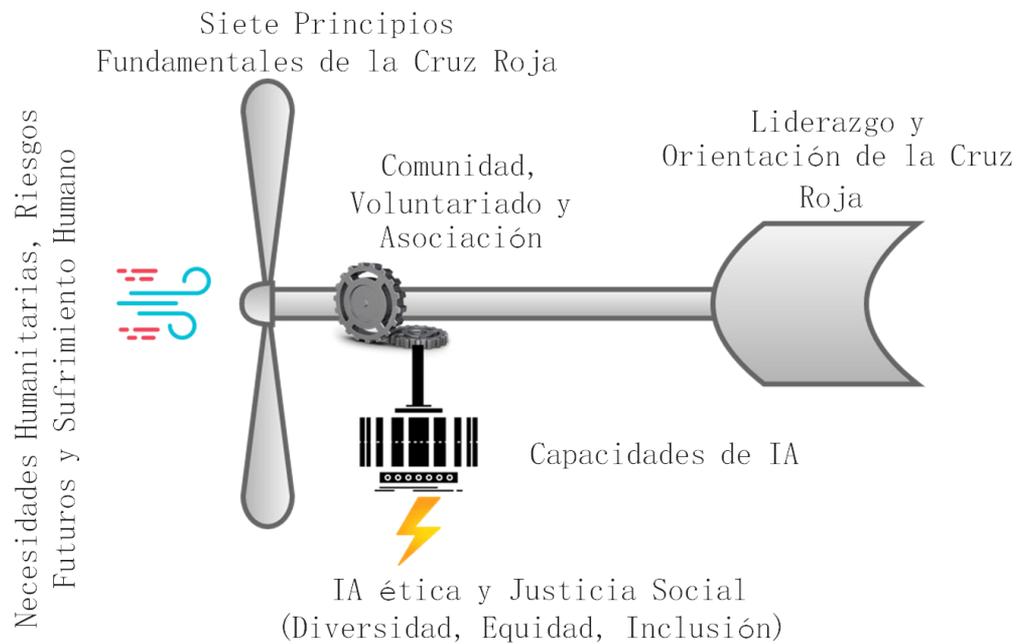


Fig.: A modelo conceptualizado para impulsar la IA de forma ética. (Fuente: <https://hyperight.com/social-justice-and-sustainability-by-leveraging-data-science-and-ai-interview-with-dr-mahendra-samarawickrama/>)

Personas, cultura y misión en la IA. ¿Qué podemos reflejar sobre AI en función de nuestra gente, cultura y misión? Discuta estos atributos y la figura ilustrada a continuación.

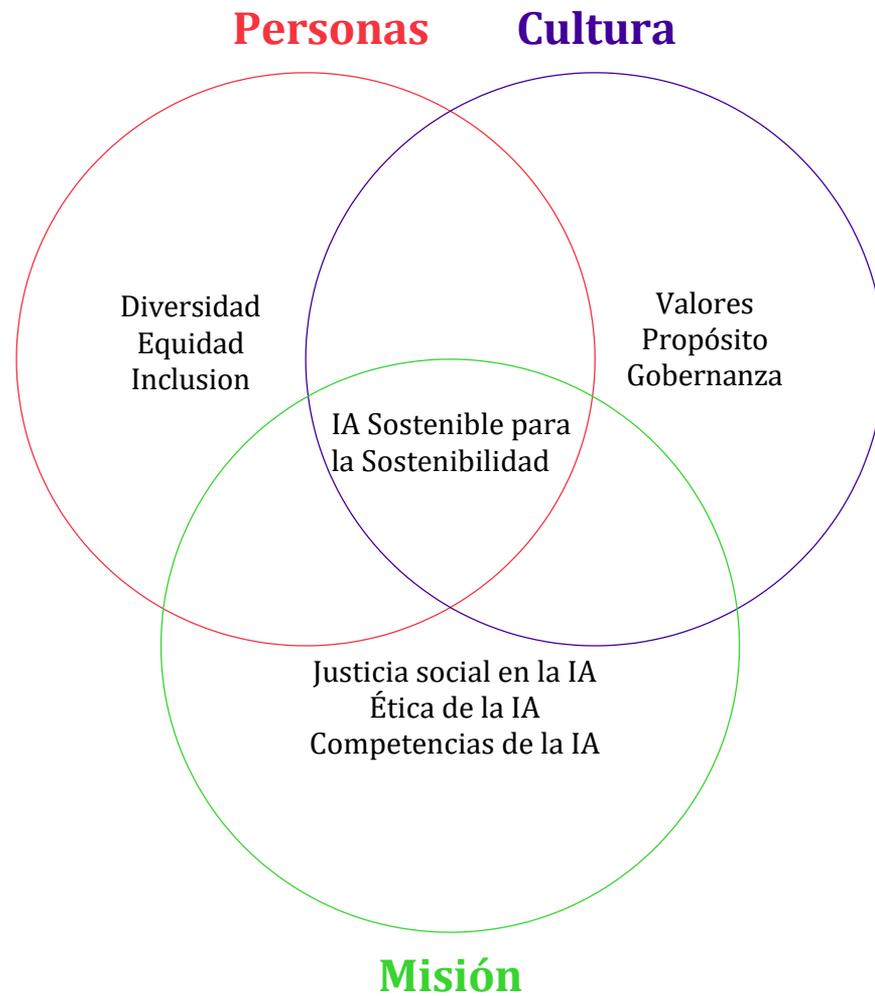


Fig.: Los atributos de las personas, la cultura y la misión en la IA para la sostenibilidad.

¿Cómo podemos convertir la Ciencia de Datos y la IA en competencias básicas? Discuta las formas de integrar la IA como parte de las competencias básicas, mientras observa el siguiente diagrama.

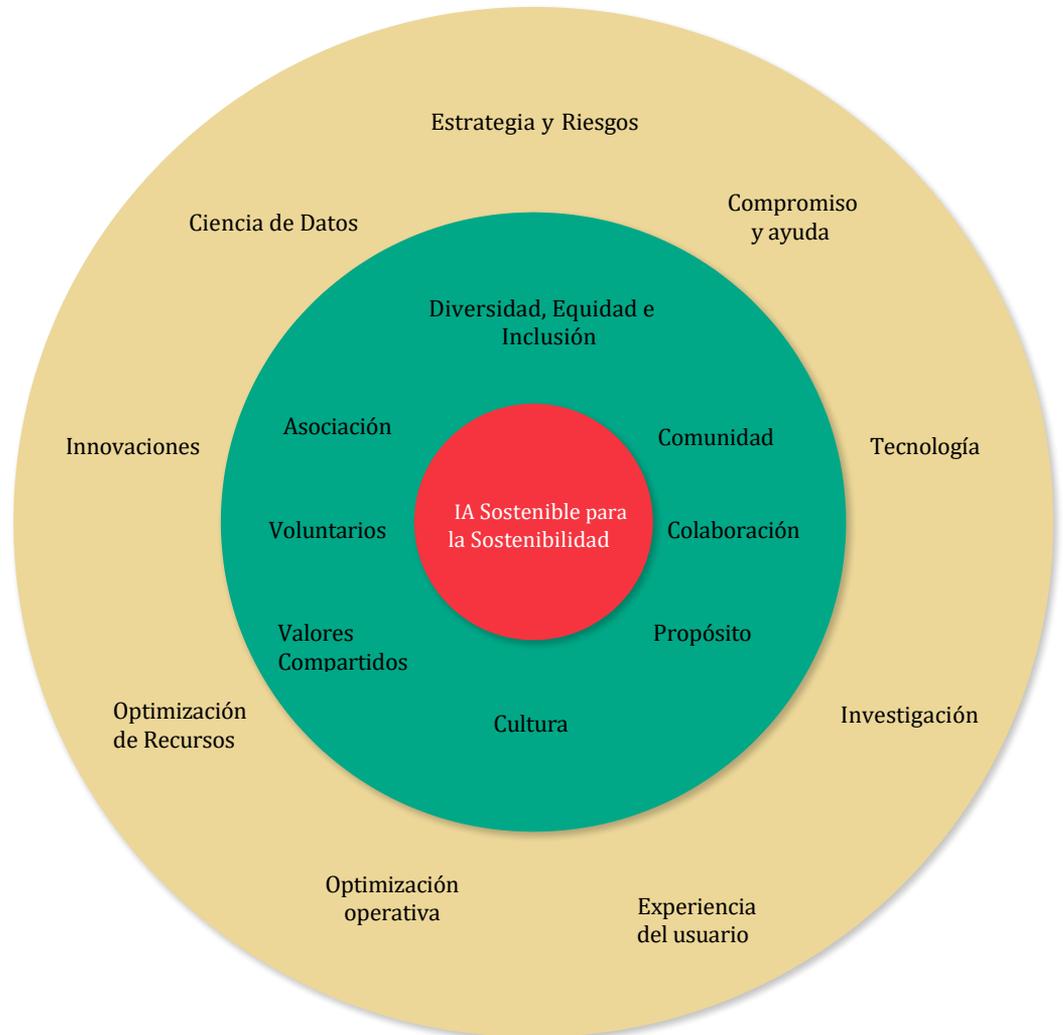
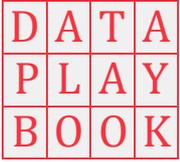


Fig.: Desarrollar la IA como competencia básica para impulsar el crecimiento de la empresa.

Parte 2: Retroalimentación

Pida a los participantes que reflexionen sobre el ejercicio y mencionen una o dos observaciones o puntos de vista.



Bono extra

Ponencias de la Cruz Roja Australiana relacionadas con la ética y la gobernanza de la IA:

Data Innovation Summit 2021 | AI for ESG Toolkit | [A Unified AI Governance Toolkit for ESG](#)

Ai4 Conference 2021 | [A Unified AI-Governance Collaborative Framework for Sustainability](#)

Crédito

Para el desarrollo del ejercicio Mahendra Smarawickrama y Paola Yela.

10 - 6 Diversidad Social, Equidad e Inclusión

Para abordar la diversidad social, la equidad y la inclusión, es necesario comprender mejor la gobernanza de las nuevas tecnologías y su papel en la sostenibilidad y la humanidad dentro de nuestra red.

- ▶ **Personas:** 2 a 8 personas
- ▶ **Tiempo:** 30 Minutos
- ▶ **Dificultad:** Media a Alta
- ▶ **Materiales Virtual:** plataforma de reunión virtual, documento/ espacio de escritura compartido
- ▶ **Materiales Presencial:** Rotafolios/pizarras, notas adhesivas, rotuladores

Ejercicio

Parte 1: Explorar y Revisar

Gobernanza de la IA y las tecnologías emergentes para la Sostenibilidad y la Humanidad

A la hora de incorporar la diversidad, la equidad y la inclusión (DEI) a los datos y las tecnologías emergentes, los líderes y colaboradores deben ser conscientes de las diferentes perspectivas sobre por qué, cómo y qué.

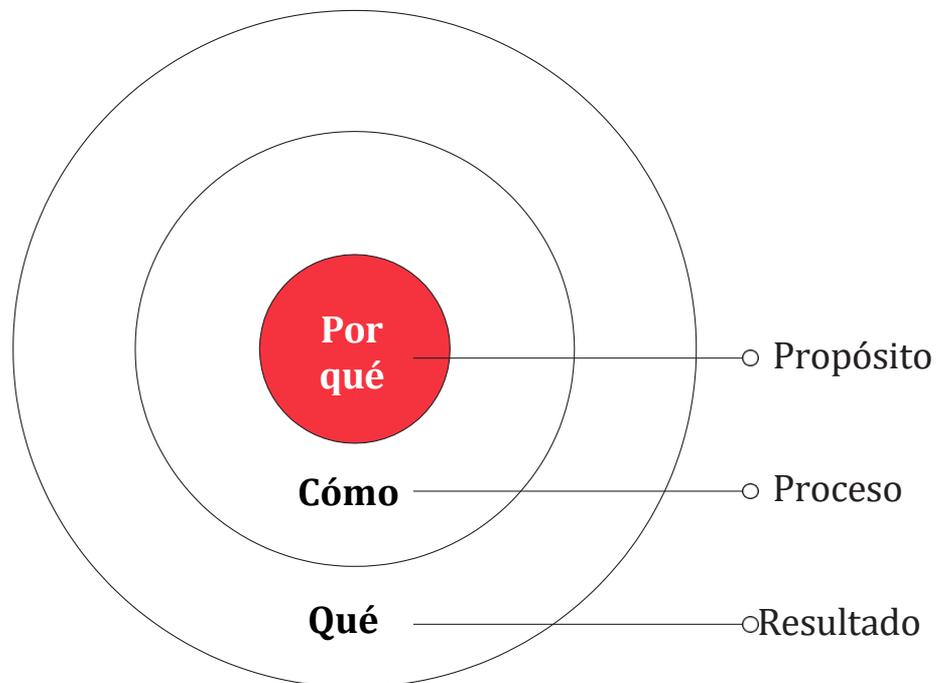
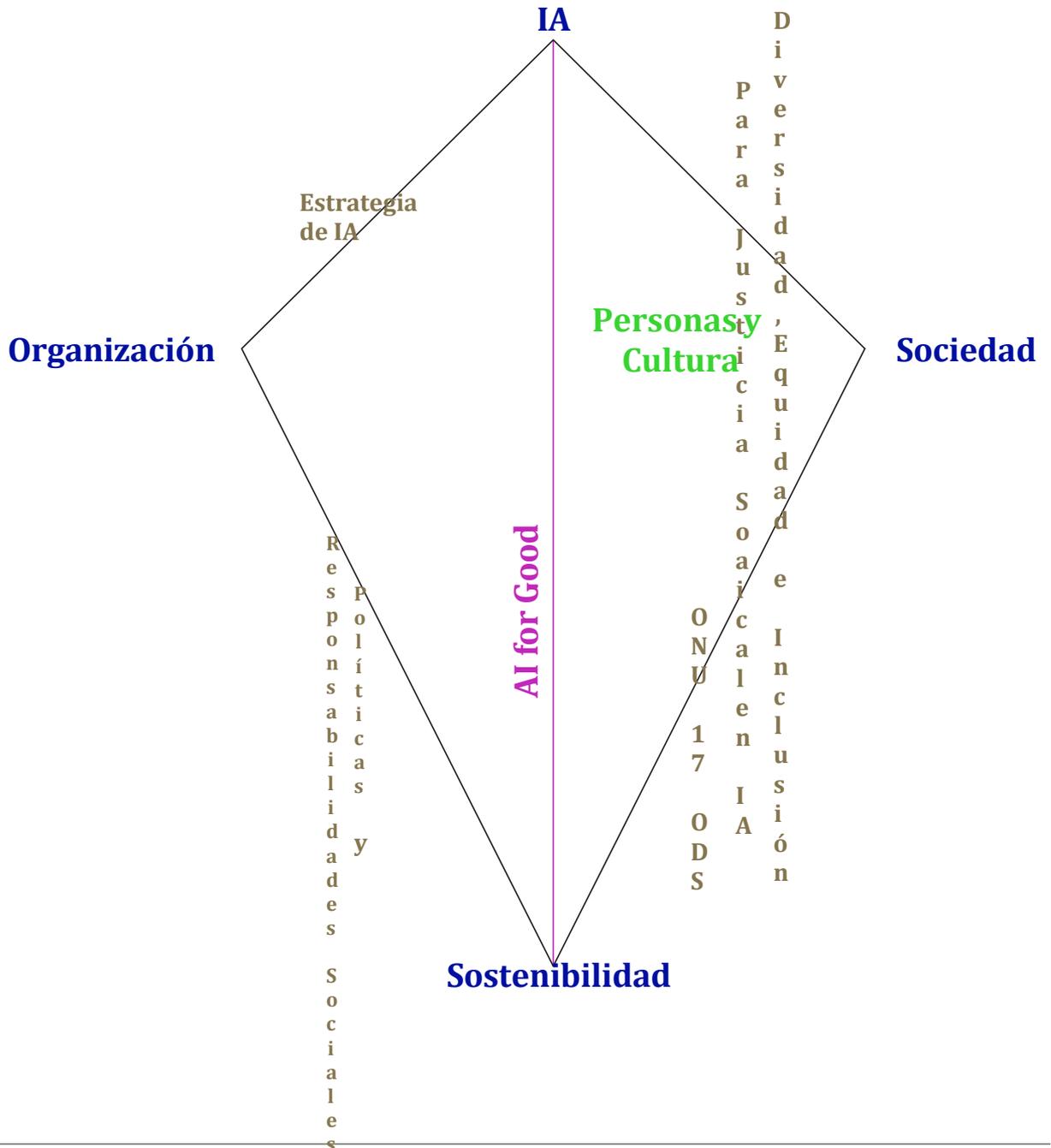


Fig: Círculo Dorado

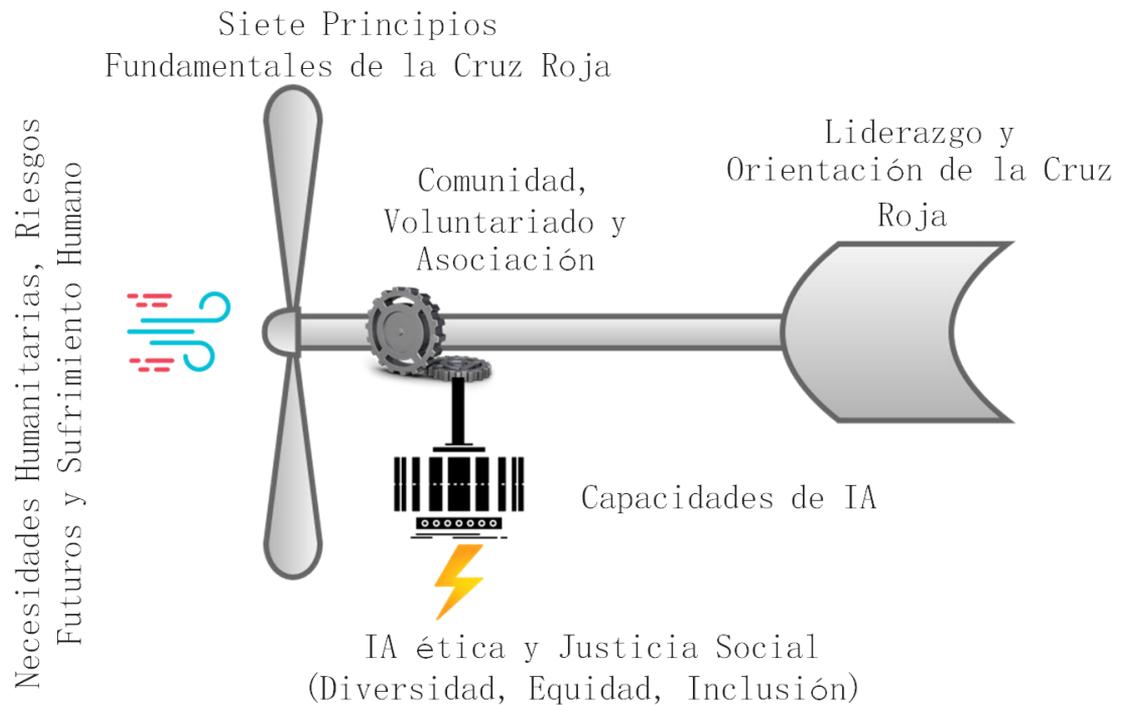
Para tener éxito en la DEI, es importante comprender la importancia de las personas, la cultura y la misión respecto a las tecnologías emergentes y mitigar el riesgo de fracaso debido a los sesgos.

Consulte las siguientes animaciones whiteboard:

¿Cuáles son los aspectos importantes de las tecnologías emergentes y la gobernanza de la IA? (Comprensión de las preguntas "Por qué"; por ejemplo, por qué los datos y la IA desempeñan un papel clave y por qué hay que tener en cuenta la gobernanza y la sostenibilidad de la IA). Pinche en el enlace para su reproducción: [The KITE Abstraction Framework for AI Governance and ESG for Sustainability.](#)



How to bring the Red Cross values to emerging technologies and AI governance? (Understanding “What” and “How” questions). Please play the Wind Turbine conceptual model whiteboard animation.



¿Puede hablarnos de cómo aporta diversidad social, equidad e inclusión a sus tecnologías emergentes o a su proyecto de IA? ¿Hasta qué punto confía en la justicia social de su proyecto? ¿Ha contribuido a la sostenibilidad (social o medioambiental)?

- ▶ Consulte el siguiente caso de éxito presentado recientemente en la Cumbre Planet:Red de la FICR
- ▶ Link de la Sesión: <https://planetredsummit.com/session/UJSGIU-1>
- ▶ Sesión de Vídeo en Streaming: <https://youtu.be/psySmdBc0IE>
- ▶ Comparta también su historia.
- ▶ Recurso adicional: Samarawickrama, M. (2022). [Keeping AI Honest. Company Director, 38\(2\), 52-53](#)

Su Sociedad	Enlace público al Caso de Éxito

Parte 2: Retroalimentación

Pida a los participantes que reflexionen sobre el ejercicio y mencionen una o dos observaciones o puntos de vista.

Bono extra

Cómo aportar diversidad social, equidad e inclusión desde el voluntariado.

- ▶ Si dispone de tiempo, podemos hablar sobre la formación de un equipo voluntario de ciencia de datos en su sociedad. Utilice para ello el Canvas de Innovación Tecnológica Emergente.
- ▶ Por favor, utilice el canvas elaborado por la Cruz Roja Australiana para conformar su diversificado equipo voluntario de ciencia de datos.
- ▶ [Canvas](#)
- ▶ [Descripción de puesto](#)
- ▶ [Equipo de voluntarios de la Cruz Roja Australiana](#)

Se puede utilizar con el Módulo **Prácticas Responsables y Protección de Datos (7)** y **Valores Humanitarios y Protección de Datos (folleto) (7 - 8)** (ejercicio).

Crédito

Para el desarrollo del ejercicio Mahendra Samarawickrama y Paola Yela

10 - 7 Mapeo de Aspectos del Trabajo en materia de Ciencia de Datos

Aclarar qué entendemos por ciencia de datos. Mapear diferentes aspectos del trabajo de la ciencia de datos, por qué necesitamos la ciencia de datos e ilustrar el tipo de productos/análisis que podemos esperar de ella. Este ejercicio permite a los equipos verificar si su conceptualización de la ciencia de datos es correcta. Al final de esta sesión, esperamos que todos tengan una visión general de la ciencia de datos, sus diferentes perspectivas, y puedan entender su rol en el espectro de trabajo de la ciencia de datos.

- ▶ **Personas:** 2 a 8 personas
- ▶ **Tiempo:** 30 Minutos
- ▶ **Dificultad:** Fácil a Media
- ▶ **Materiales Virtual:** plataforma de reunión virtual, documento/ espacio de escritura compartido
- ▶ **Materiales Presencial:** Rotafolios/pizarras, notas adhesivas, rotuladores

Ejercicio

Parte 1: Explorar and Revisar

Por favor, analice cómo se puede conformar un equipo de ciencia de datos tomando como referencia el siguiente diagrama y dónde se ubican mejor sus funciones en el espectro.

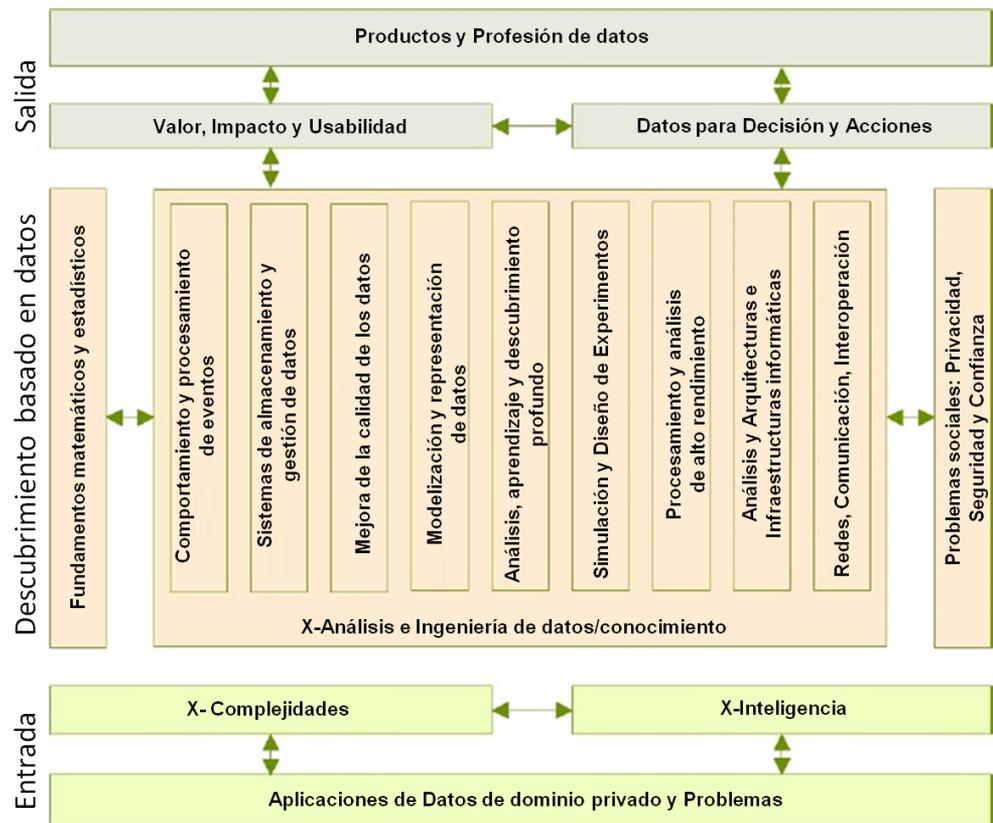
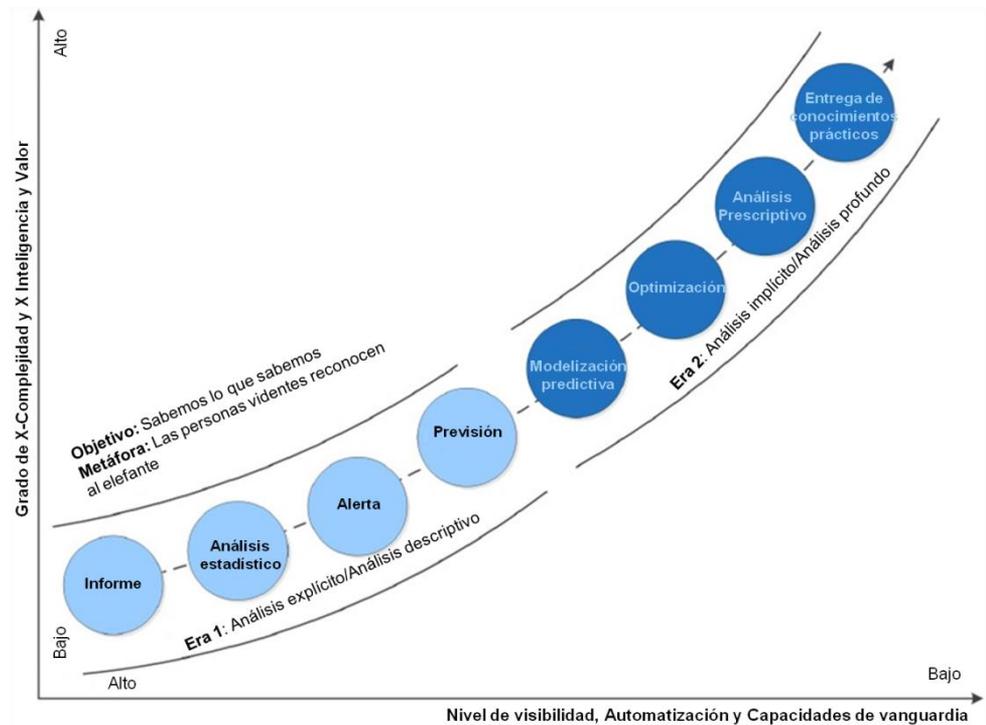


Fig. 6 Mapa conceptual Ciencia de Datos

(i) Tomado de aquí: Fuente: <https://arxiv.org/abs/2007.03606>

Analice los principales componentes de la ciencia de datos que se ilustran en el siguiente diagrama y reflexione sobre si se ajustan a las necesidades de su Sociedad Nacional a la hora de formar su equipo de datos. ([Data Science: A Comprehensive Overview](#))



Parte 2: Retroalimentación

Pida a los participantes que reflexionen sobre el ejercicio y mencionen una o dos observaciones o puntos de vista.

Bono extra

El facilitador puede continuar este debate refiriéndose al ejercicio "¿Por qué necesitamos la ciencia de datos?" del Módulo 10, relacionando los puntos clave de esta sesión con la exploración de las capacidades de un equipo de ciencia de datos.

Crédito

Para el desarrollo del ejercicio Mahendra Samarawickrama y Paola Yela

10 - 8 ¿Por qué necesitamos la Ciencia de Datos?

Clarificar lo que entendemos por ciencia de datos. Mapear diferentes aspectos del trabajo de la ciencia de datos, por qué necesitamos la ciencia de datos e ilustrar el tipo de productos/análisis que podemos esperar de ella. Este ejercicio explora la necesidad de establecer un equipo de ciencia de datos y ayuda a descubrir algunos aspectos clave que es necesario conocer cuando se empieza a trabajar con equipos de ciencia de datos en la CRMLR.

- ▶ **Personas:** 2 a 8 personas
- ▶ **Tiempo:** 45 - 60 Minutos
- ▶ **Dificultad:** Media a Alta
- ▶ **Materiales Presencial:** Rotafolios y rotuladores
- ▶ **Materiales Virtual:** plataforma de reunión virtual, documento/ espacio de escritura compartido

Ejercicio

Parte 1: Explorar y Revisar

Estudio de caso: El equipo de dirección de la sociedad de la Cruz Roja ABC decidió crear un equipo de ciencia de datos que pueda impulsar la ciencia de datos y el análisis para impulsar la actividad.

En relación con el caso, permita que los grupos debatan una serie de preguntas que puedan ayudarles a explorar la necesidad de conformar un equipo de ciencia de datos y si eso les acerca un paso más a la identificación del uso del aprendizaje automático (machine learning) en el ámbito de la ciencia de datos.

Preguntas que se utilizarán para el debate:

- 1 ¿Qué es la ciencia de datos?
 - (nota para el facilitador - las respuestas pueden incluir: ciencia de datos = estadística + informática + computación + comunicación + sociología + gestión | datos + medio ambiente + pensamiento)
- 2 ¿Por qué los niveles directivos necesitan un equipo de ciencia de datos?
 - (nota para el facilitador - las respuestas pueden incluir: ayudar a tomar decisiones basadas en datos, personalizar la experiencia del usuario, ayudar a hacer predicciones y gestionar riesgos, diversificar los servicios y la estrategia, etc.)
- 3 ¿Cuáles son las capacidades actuales que debe tomar en cuenta el equipo directivo?
 - (nota para el facilitador - las respuestas pueden incluir: plataformas, competencias de las personas, herramientas que se utilizan, la cultura de datos, etc.)
- 4 ¿En qué inversiones debería estar dispuesto a destinar fondos el equipo directivo?
 - (nota para el facilitador- las respuestas pueden incluir: formación del personal, adquisición de tecnología como herramientas y plataformas, etc.)
- 5 ¿Qué competencias incluye el equipo directivo en la nueva descripción del puesto?

Bono extra

Para profundizar en la integración del aprendizaje automático en los proyectos de ciencia de datos, los grupos tienen la opción de responder a las siguientes preguntas:

- 1 ¿Ha oído hablar de los algoritmos de aprendizaje automático?
- 2 ¿Puede explicárselos a una persona sin conocimientos técnicos?
- 3 ¿Qué significa garbage in garbage out (basura entra, basura sale) en la modelización de Ciencia de Datos?
- 4 ¿Qué es una hipótesis en la modelización de ciencia de datos?
- 5 ¿Puede explicar el sesgo y la varianza en la modelización de ciencia de datos utilizando la hipótesis?

Posibilidad de que el facilitador consulte el material adicional Rol de un equipo de voluntarios en materia de Ciencia de Datos y Cómo sería la descripción del puesto de un científico de datos, que se proporcionan en el módulo 10.

Crédito

Para el desarrollo del ejercicio Mahendra Samarawickrama y Paola Yela, Editado por Melissa El Hamouch

Recurso extra: Eggleton, M., 2022. Award-winner warns of the failures of artificial intelligence. [The Australian Financial Review - Special Report of Reimagination 2022](#), pp.S4-S5.

10 - 9 Decisiones de Ayuda en Contextos Específicos

Cuando se dispone de varias opciones de procesos de ciencia de datos, no siempre es fácil identificar cuál es la mejor opción para un proyecto determinado. Este ejercicio ayudará a los equipos a debatir las opciones disponibles en función de sus recursos y capacidades.

- ▶ **Personas:** 2 a 8 personas
- ▶ **Tiempo:** 45 – 60 Minutos
- ▶ **Dificultad:** Alta
- ▶ **Materiales Virtual:** plataforma de reunión virtual, documento/ espacio de escritura compartido
- ▶ **Materiales Presencial:** Rotafolios/pizarras, notas adhesivas, rotuladores

Ejercicio

Parte 1: Explorar y Revisar

Este ejercicio permite a los equipos responder a preguntas sobre sus recursos, capacidades y tiempo para comprender cuál es la opción que más les conviene seleccionar para un objetivo concreto. El facilitador puede proponer una tarea de proyecto teórica o debatir una que proponga el equipo.

Nota: Pida a los equipos que escriban diferentes opciones que puedan cotejar con los recursos de que disponen y la situación en la que se encuentran.

A continuación, el facilitador puede guiar al equipo en la identificación de la opción que mejor se adapte a su situación. Se puede hacer una lista de preguntas para ayudar a los grupos y equipos a reflexionar.

Las preguntas pueden ser las siguientes:

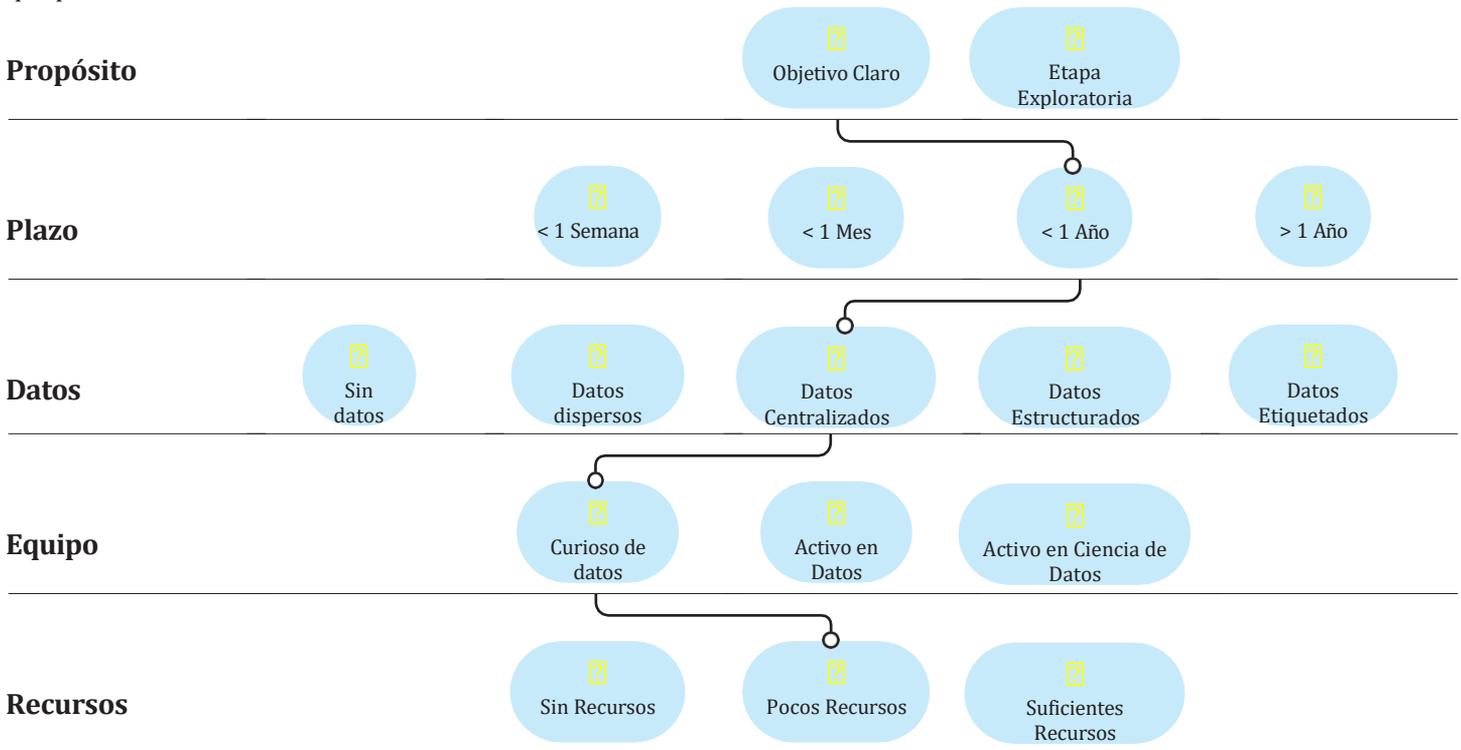
- ▶ ¿De cuántos recursos dispone?
- ▶ ¿De cuánto tiempo dispone?
- ▶ ¿Cuánto tiempo llevará trabajar en el proyecto?
- ▶ ¿Qué posibles limitaciones podrían surgir?
- ▶ ¿Cuál es el presupuesto asignado al proyecto?
- ▶ ¿De cuánto personal se dispone para trabajar en el proyecto?
- ▶ ¿Qué competencias necesita el equipo?
- ▶ ¿Cuál es el entorno en el que se desarrollará el trabajo (entorno de emergencia o no)?
- ▶ ¿De qué datos dispone el equipo (datos históricos, datos secundarios, ausencia de datos, etc.) y qué datos necesita tener?

El facilitador tiene la opción de presentar al equipo los siguientes casos como punto de partida para el debate. A continuación, puede remitirse al conjunto de preguntas presentadas anteriormente para obtener un mayor conocimiento del equipo y evaluar lo que se necesita y lo que se tiene disponible. Esto les ayudará a decidir qué proceso y recursos de la ciencia de datos serán más eficaces.

CASO A:

La Sociedad Nacional XYZ dispone de un repositorio de evaluaciones comunitarias de riesgos realizadas en varias regiones de su país. Se trata de documentos no estructurados de contenido similar pero con diferentes formatos.

Al equipo de gestión de riesgos de la Sociedad Nacional le gustaría obtener información clave de esos documentos, pero no tienen capacidad interna para procesar todos los documentos. No es una tarea urgente, pero disponen de algunos recursos que podrían invertirse en esta tarea.



El equipo de la Sociedad Nacional XYZ decidió hacer unas prácticas de 4 meses en una institución técnica. ¿Haría lo mismo su equipo? ¿Qué opción habría preferido?

CASO B:

Seleccione su ruta en función de la situación:

Propósito

Objetivo Claro

Etapa
Exploratoria**Plazo**< 1
Semana

< 1 Mes

< 1 Año

> 1 Año

DatosSin
datosDatos
dispersosDatos
centralizadosDatos
estructuradosDatos
etiquetados**Equipo**Curioso de
DatosActivo en
DatosActivo en Ciencia de
Datos**Recursos**

Sin Recursos

Pocos Recursos

Suficientes
Recursos

¿Qué opción consideraría? ¿Qué otras opciones propondría para esta situación?

Pasantías

Asociaciones de investigación

Consultorías

Hackathons

Datos internos

Proyecto de recolección

Proyecto DS interno

Ayuda al movimiento

Asociaciones tecnológicas

Proyecto de etiquetado interno

Asociaciones de datos

Implicar voluntarios

Grupos de trabajo

Parte 2: Retroalimentación

Pida a los participantes que reflexionen sobre el ejercicio y mencionen una o dos observaciones o puntos de vista.

Bono extra

Para más información, los equipos pueden consultar [AI Suitability Framework](#) que proporciona un conjunto de preguntas que deben plantearse a lo largo del desarrollo de un proyecto de ciencia de datos. Aporta suficientes detalles e intenta mantener una perspectiva práctica. El gráfico representado más arriba puede considerarse una primera aproximación al mismo.

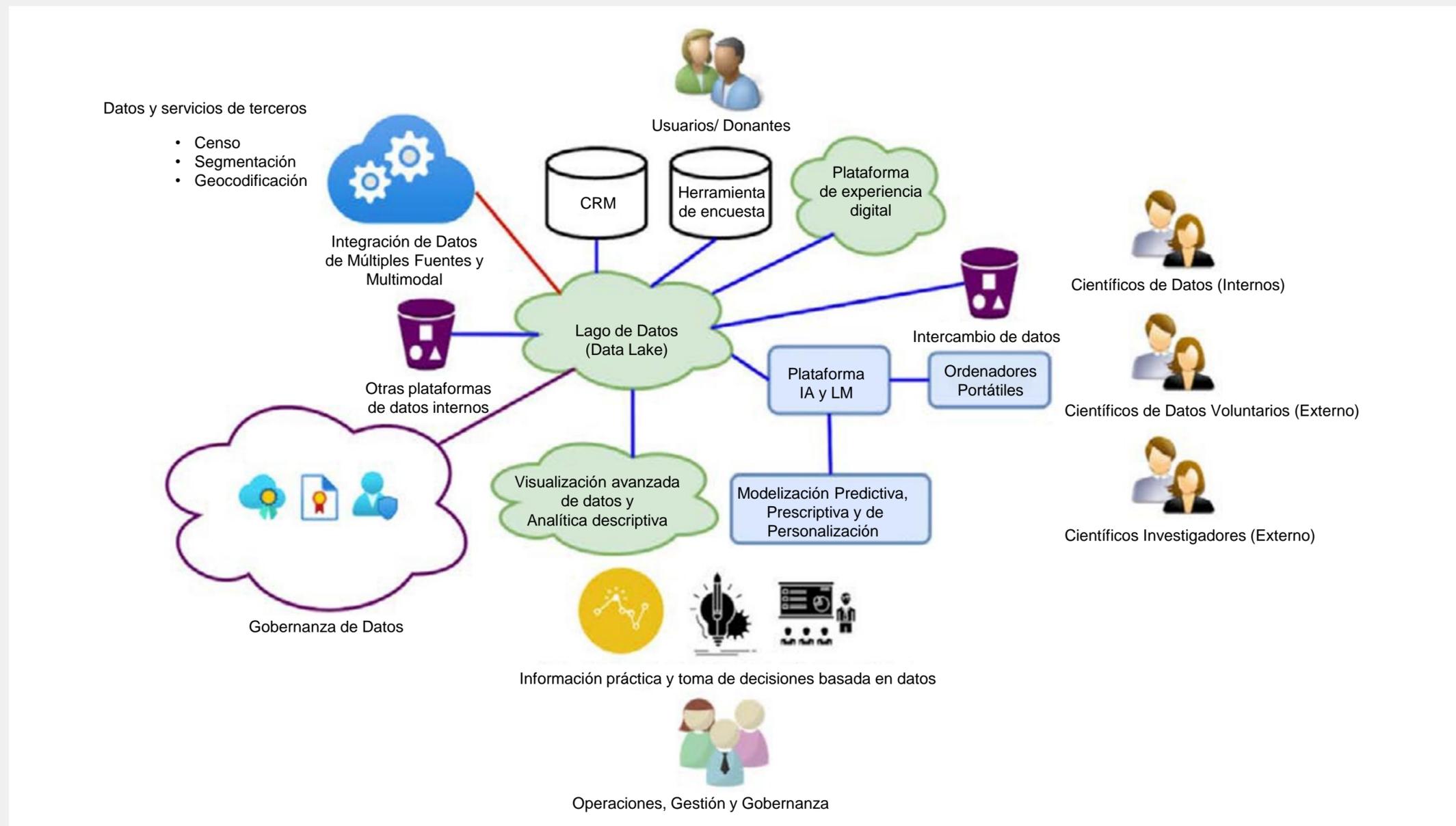
Crédito

Para el desarrollo del ejercicio Paola Yela y Melissa El Hamouch

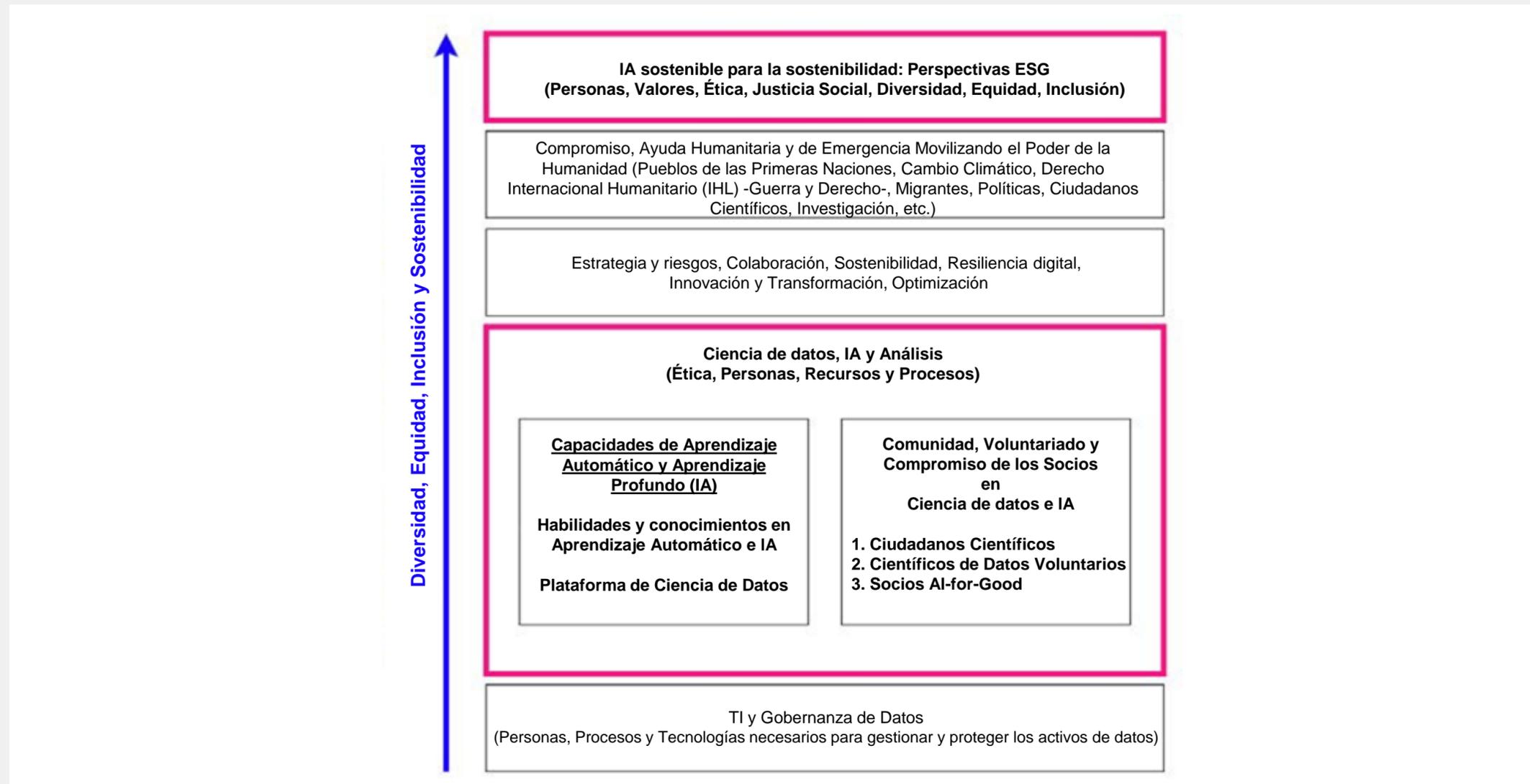
10 - 10 Visión general de la Plataforma y los Procesos típicos de la Ciencia de Datos

Créditos: Mahendra Samarawickrama para el Data Playbook de la FICR

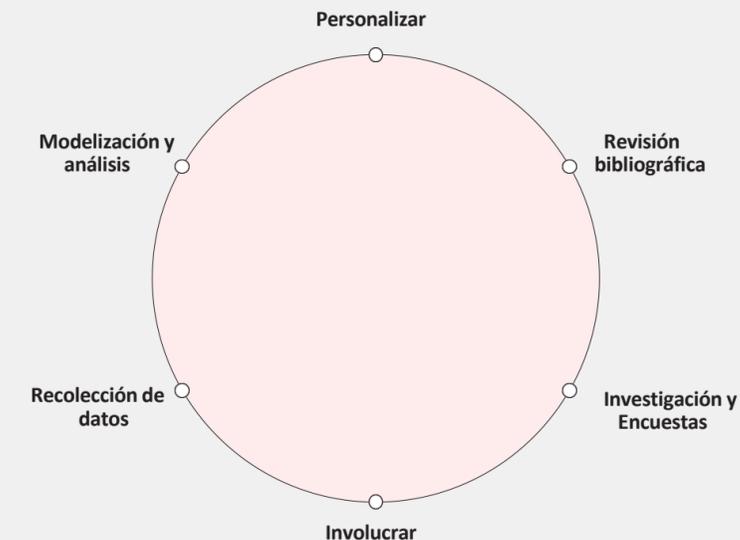
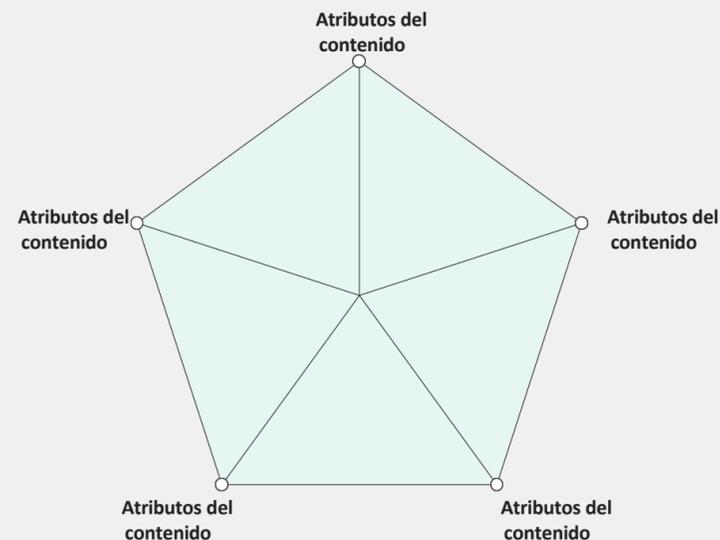
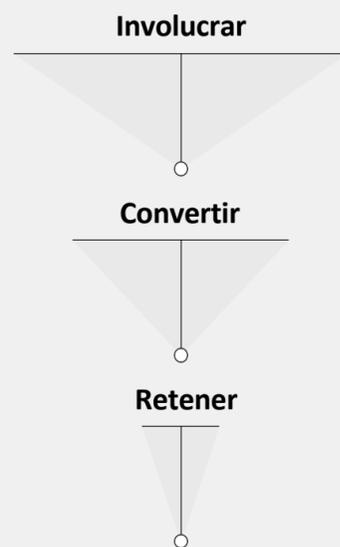
Esquema típico de una Plataforma de Ciencia de Datos



Proceso típico de Ciencia de Datos centrado en objetivos Humanitarios



Ciencia de Datos para Personalización



Compromiso y apoyo

CX

Adquisición y Fidelización

Digital

Marketing

Finanzas

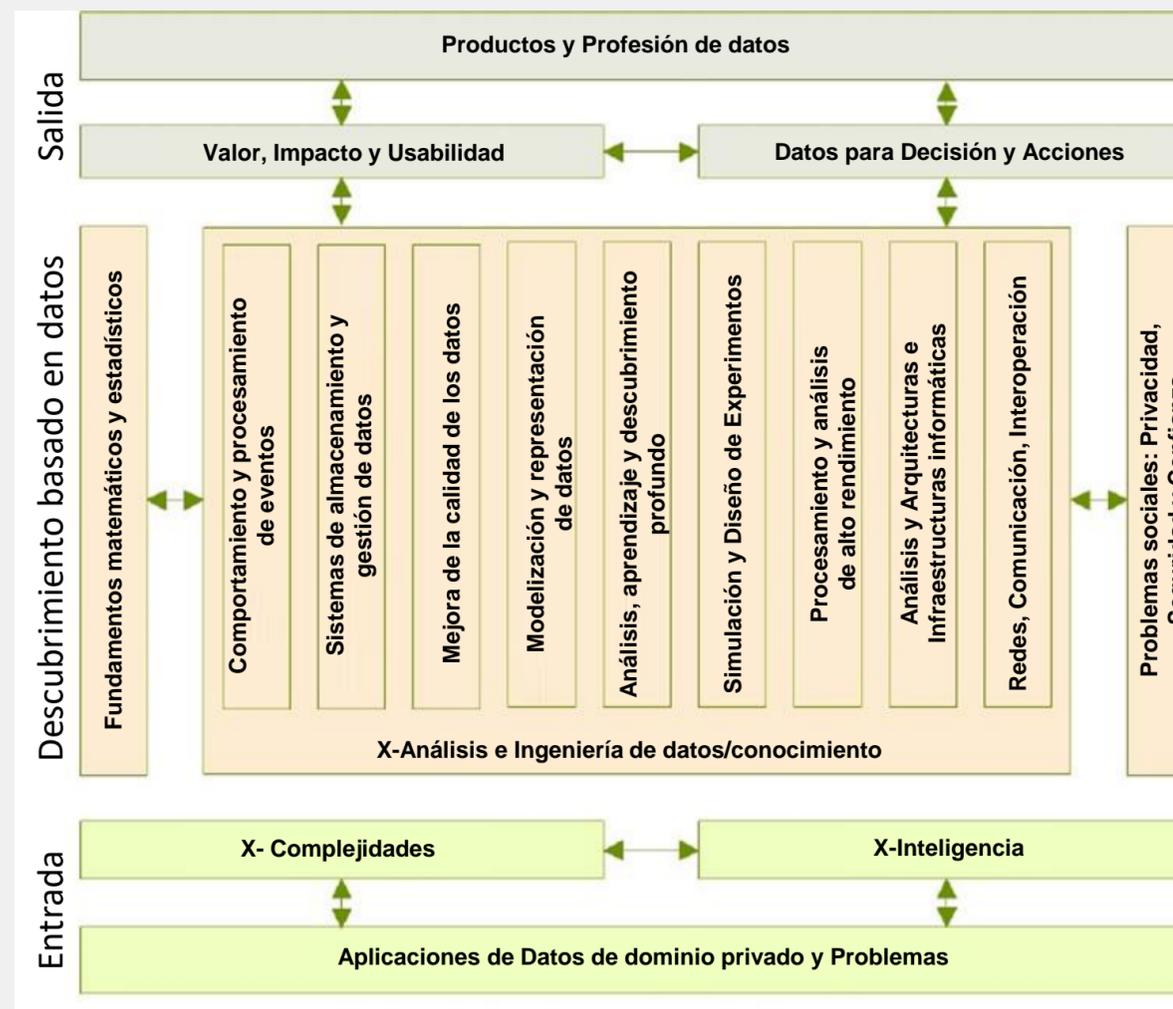
Data Insight (Conocimiento de los datos) como un Servicio

Ciencia de datos (p.e., Personalización, Predicción y Análisis avanzado)

Arquitectura de datos (p.e., infraestructura)

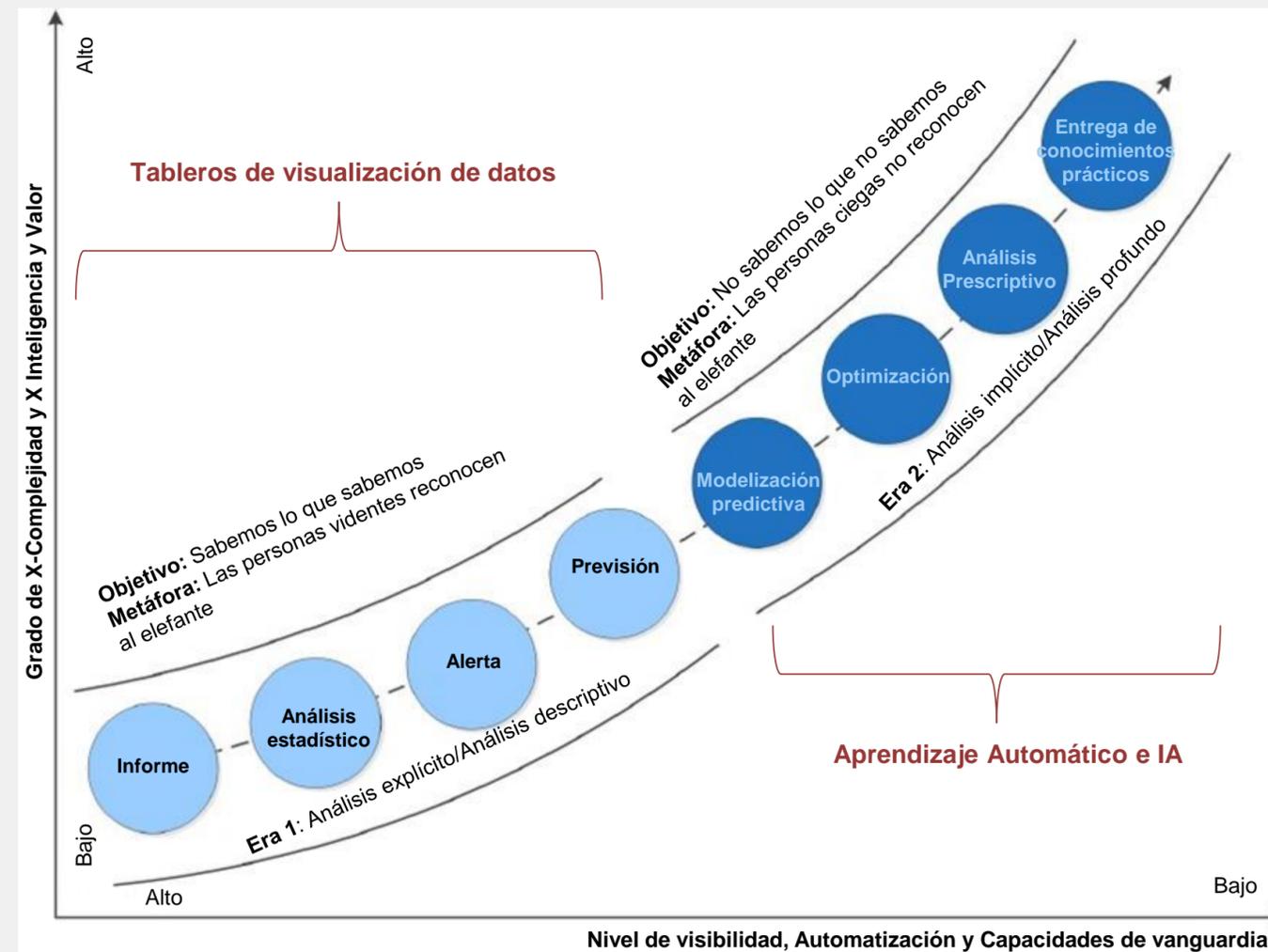
IT (p.e., Seguridad de los datos)

Mapa conceptual de Ciencia de Datos



(i) Fuente: Longbing Cao. 2017. Ciencia de datos: una descripción completa. ACM Comput. Surv. 50, 3, Artículo 43 (Junio 2017), página 42. DOI: <https://doi.org/10.1145/3076253>

Hoja de ruta de la Ciencia de Datos (Creación de valor)



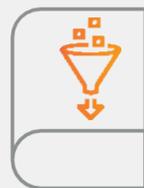
(i) Fuente: Longbing Cao. 2017. Ciencia de datos: una descripción completa. ACM Comput. Surv. 50, 3, Artículo 43 (Junio 2017), página 42. DOI: <https://doi.org/10.1145/3076253>

Plataforma Típica de Ciencia de Datos

1. Data Lake



AWS Glue



AWS Glue Catalog



Athena



S3 Bucket

2. Aprendizaje Automático y Plataforma de IA



Servicios IA:

Amazon Rekognition • Amazon Polly • Amazon Lex

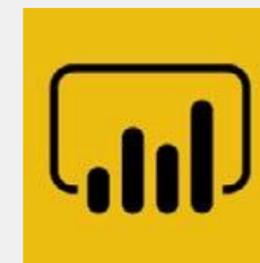
Plataformas de IA:

Amazon Machine Learning • Amazon EMR
• Spark & Spark ML

Motores de IA:

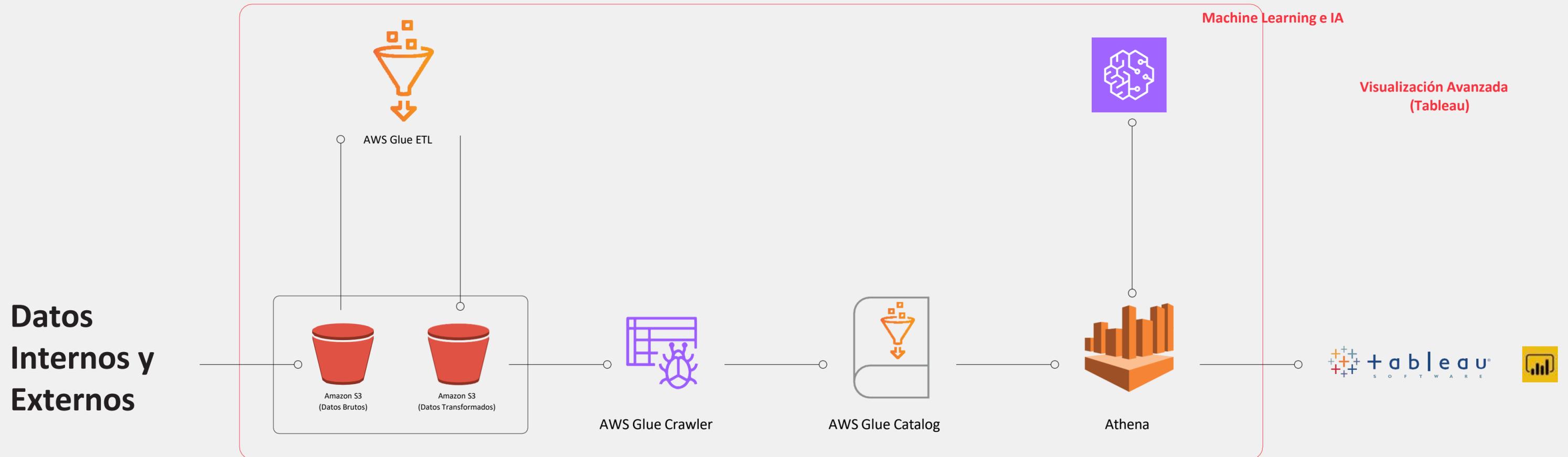
Apache MXNet • TensorFlow • Caffe •
Torch • Theano • CNTK • Keras

3. Marco de visualización

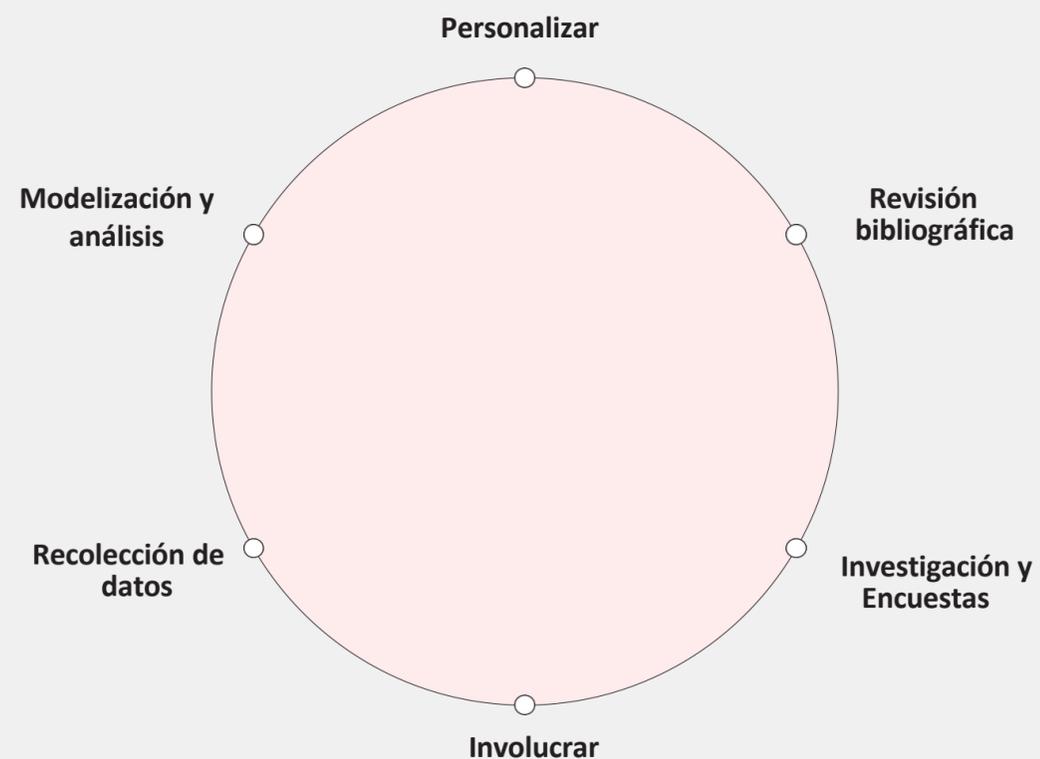
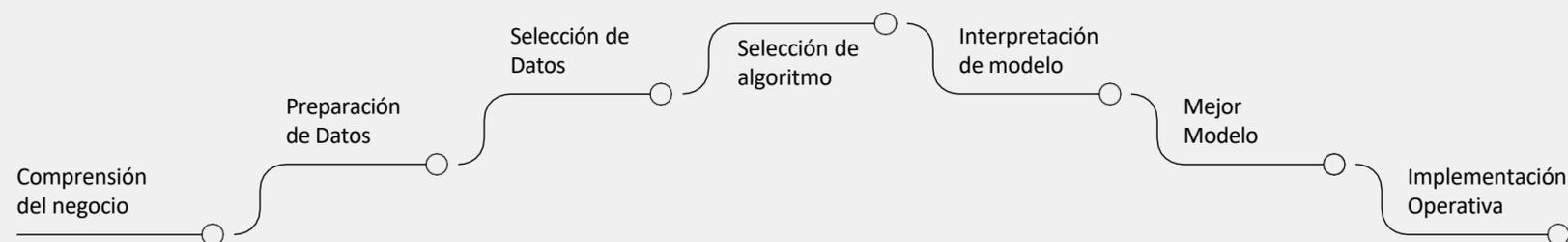


Ejemplo de una Plataforma Típica de Ciencia de Datos (sólo con fines ilustrativos)

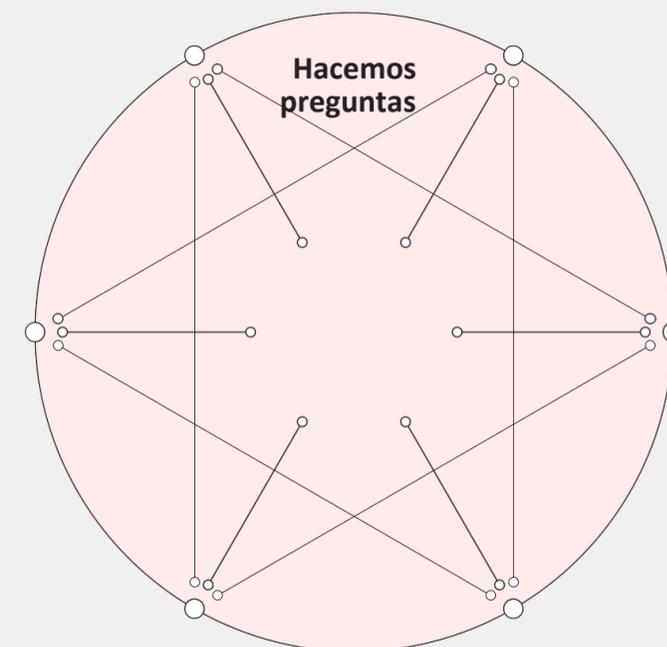
Data Lake + Machine Learning + IA + Visualización (Simple, de bajo coste y diseñado a medida)



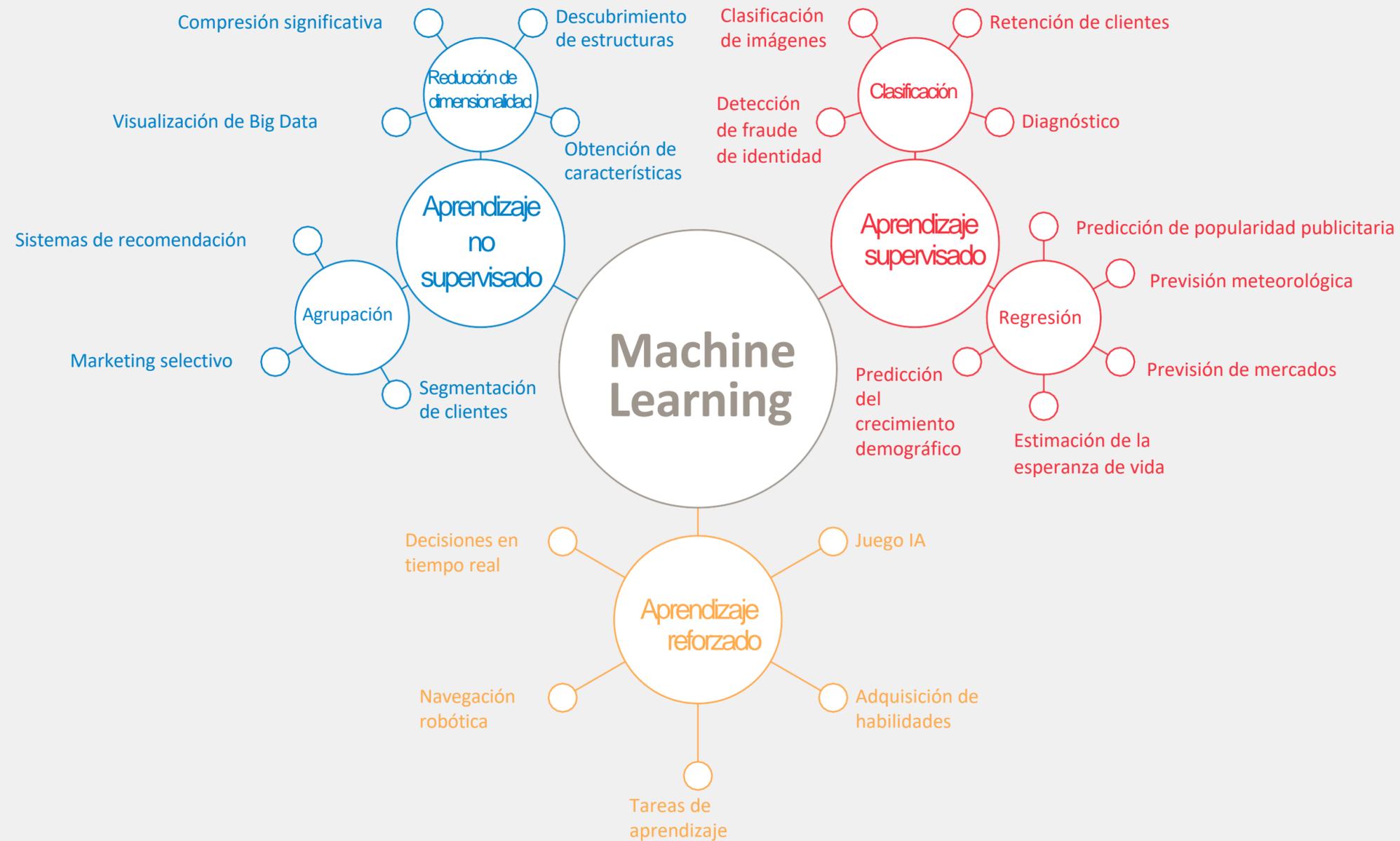
Proceso de Ciencia de Datos (Proceso Iterativo e Interactivo)



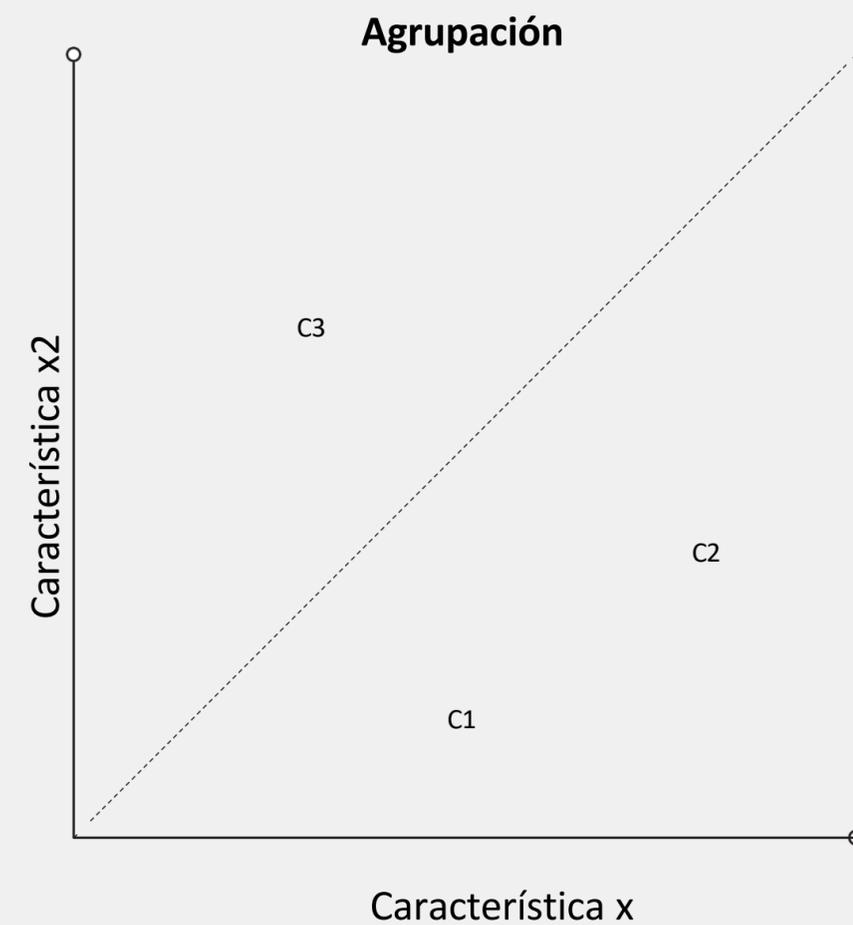
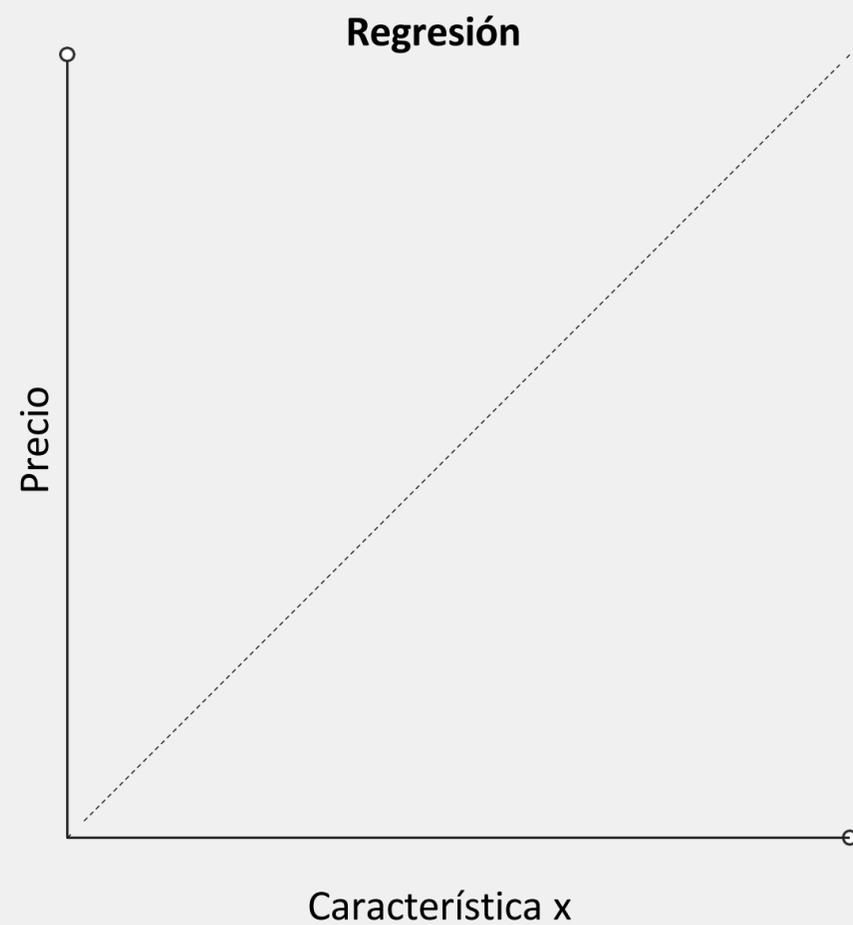
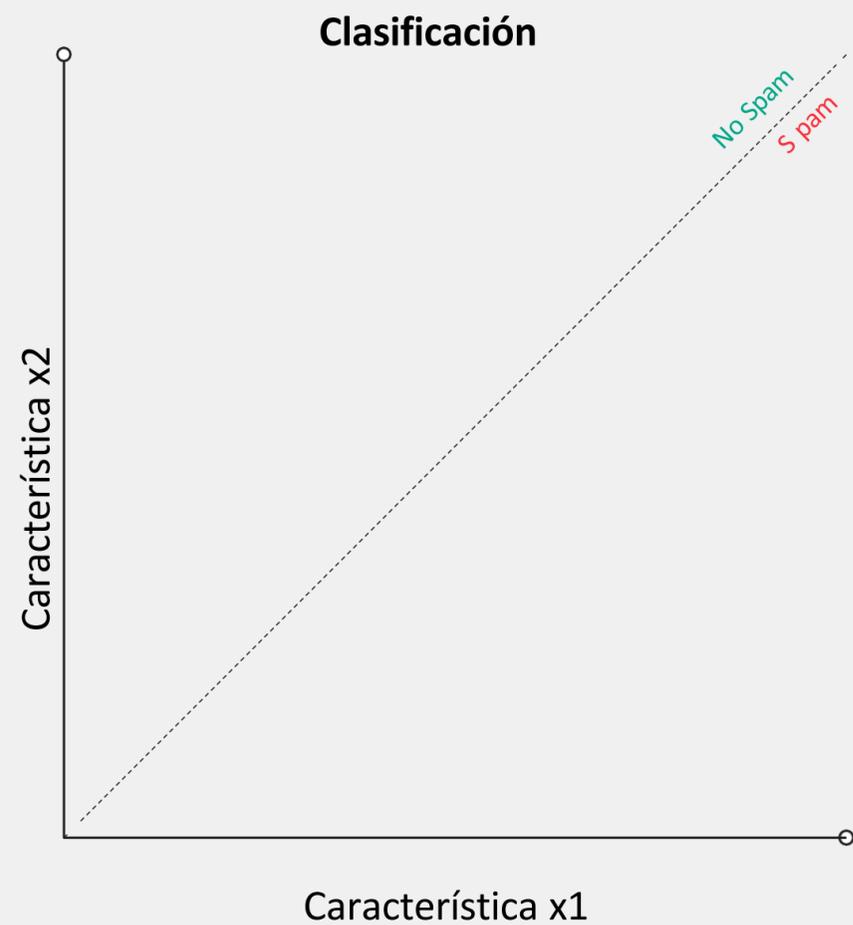
•



Marco de referencia del Aprendizaje Automático (Machine Learning)



Machine Learning: Descripción general de alto nivel



¡Gracias!

10 - 11 **Cómo implementar un modelo de Ciencia de Datos**

Este ejercicio permitirá a los equipos adquirir experiencia práctica en aprendizaje automático. Al permitir que los equipos comprendan mejor cómo se implementa un modelo de ciencia de datos, podrán evaluar mejor qué se puede integrar en su trabajo.

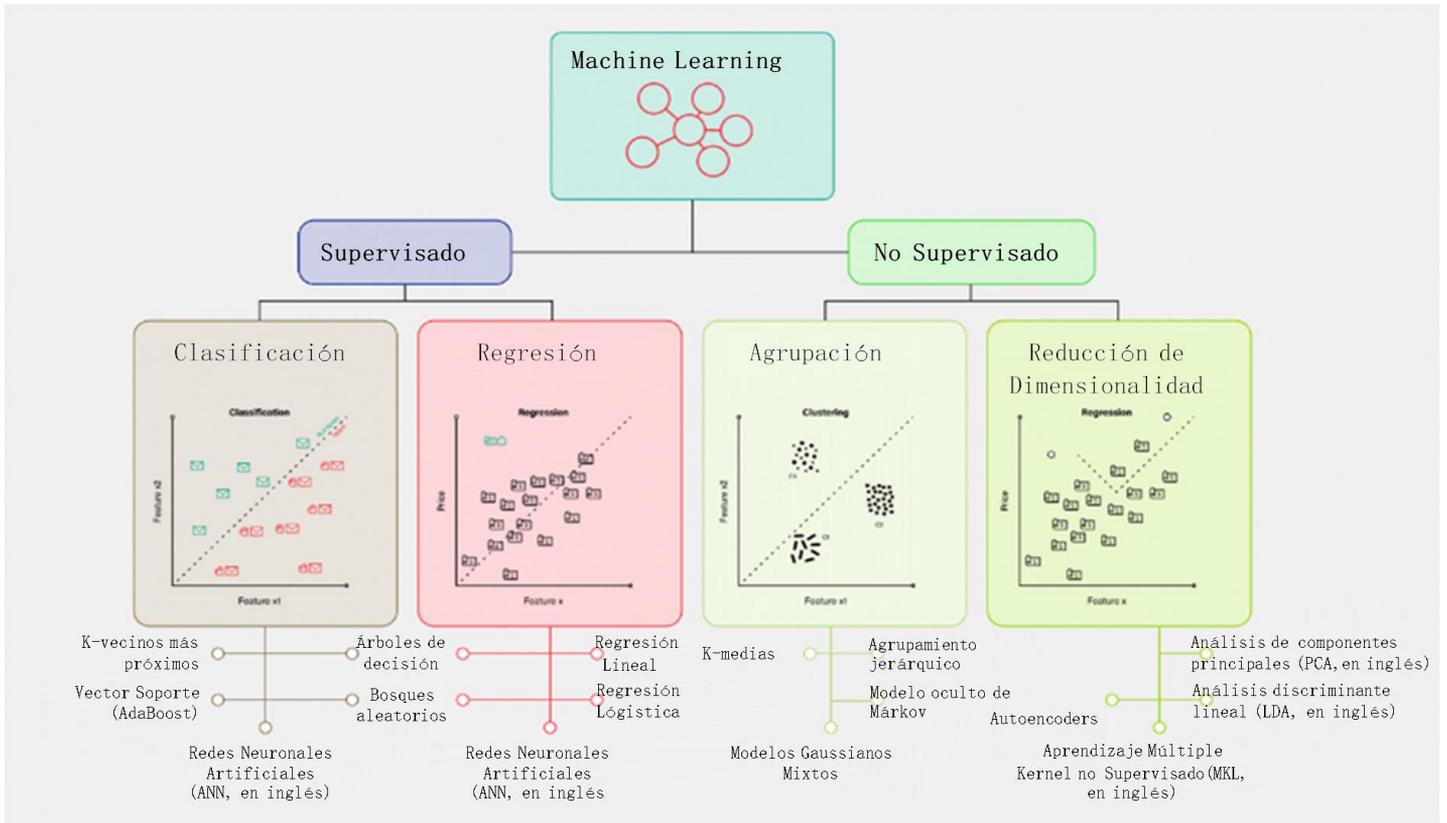
- ▶ **Personas:** 2 a 8 personas
- ▶ **Tiempo:** 45 Minutos
- ▶ **Dificultad:** Media a Alta
- ▶ **Materiales Virtual:** plataforma de reunión virtual, documento/ espacio de escritura compartido
- ▶ **Materiales Presencial:** Rotafolios/pizarras, notas adhesivas, rotuladores

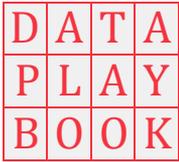
Ejercicio

Parte 1: Explorar y Revisar

Ejercicio Práctico de Aprendizaje Automático (Cómo implementar un modelo de ciencia de datos)

- 1 Hemos preparado un Jupyter Notebook para ilustrar un sencillo modelo de clasificación de aprendizaje automático. [Cómo configurar Google Colab](#). Observe qué es un modelo de clasificación en el siguiente diagrama.





- 2 Vamos a desarrollar un modelo simple de churn.
- 3 Como se trata de un modelo de clasificación binaria, necesitamos conocer las etiquetas de los donantes antes de modelizar. En este caso, necesitamos saber quiénes eran donantes activos y quiénes no. Entrenamos este modelo basándonos en datos históricos.
- 4 Obsérvese lo siguiente
 - Etapa de preparación de los datos
 - Etapa de muestreo
 - Etapa de análisis exploratorio de los datos
 - División de los datos en entrenamiento y validación
 - Entrenamiento del modelo (Obsérvese el modelo de clasificación LightGBM)
 - tapa de validación (Obsérvese qué medida se utilizó para validar la precisión del modelo)
 - Análisis del modelo
- 5 Puede continuar leyendo sobre este tipo de modelización de aprendizaje automático:
 - <https://youtu.be/8b1JEDvenQU>
 - <https://www.youtube.com/watch?v=GM3CDQfQ4sw>
 - <https://www.youtube.com/watch?v=4jRBRDbJemM>

Parte 2: Retroalimentación

Pida a los participantes que reflexionen sobre el ejercicio y mencionen una o dos observaciones o puntos de vista.

Bono extra

Para más información sobre el tema, puede consultar los siguientes enlaces

- 1 <https://unstats.un.org/unsd/undataforum/blog/KITE-an-abstraction-framework-for-reducing-complexity-in-ai-governance/>
- 2 <https://hyperight.com/social-justice-and-sustainability-by-leveraging-data-science-and-ai-interview-with-dr-mahendra-samarawickrama/>

Crédito

Para el desarrollo del ejercicio Mahendra Smarawickrama y Paola Yela.

10 - 12 **Análisis de Churn de la Cruz Roja: Jupyter Notebook**

[Recursos para descargar](#)

[Recursos para descargar](#) (Data Model CSV)

[Recursos para descargar](#) (Data Model SQL)