



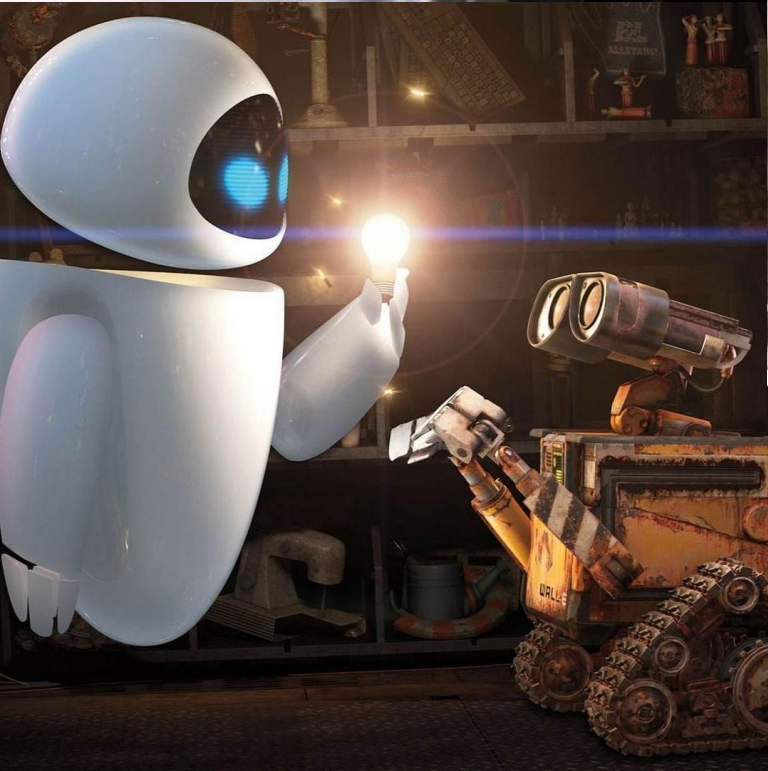
Conceptos Básicos

Python (Machine Learning + Big Data)

Centro de Servicios y Gestión Empresarial
SENA Regional Antioquia



www.sena.edu.co

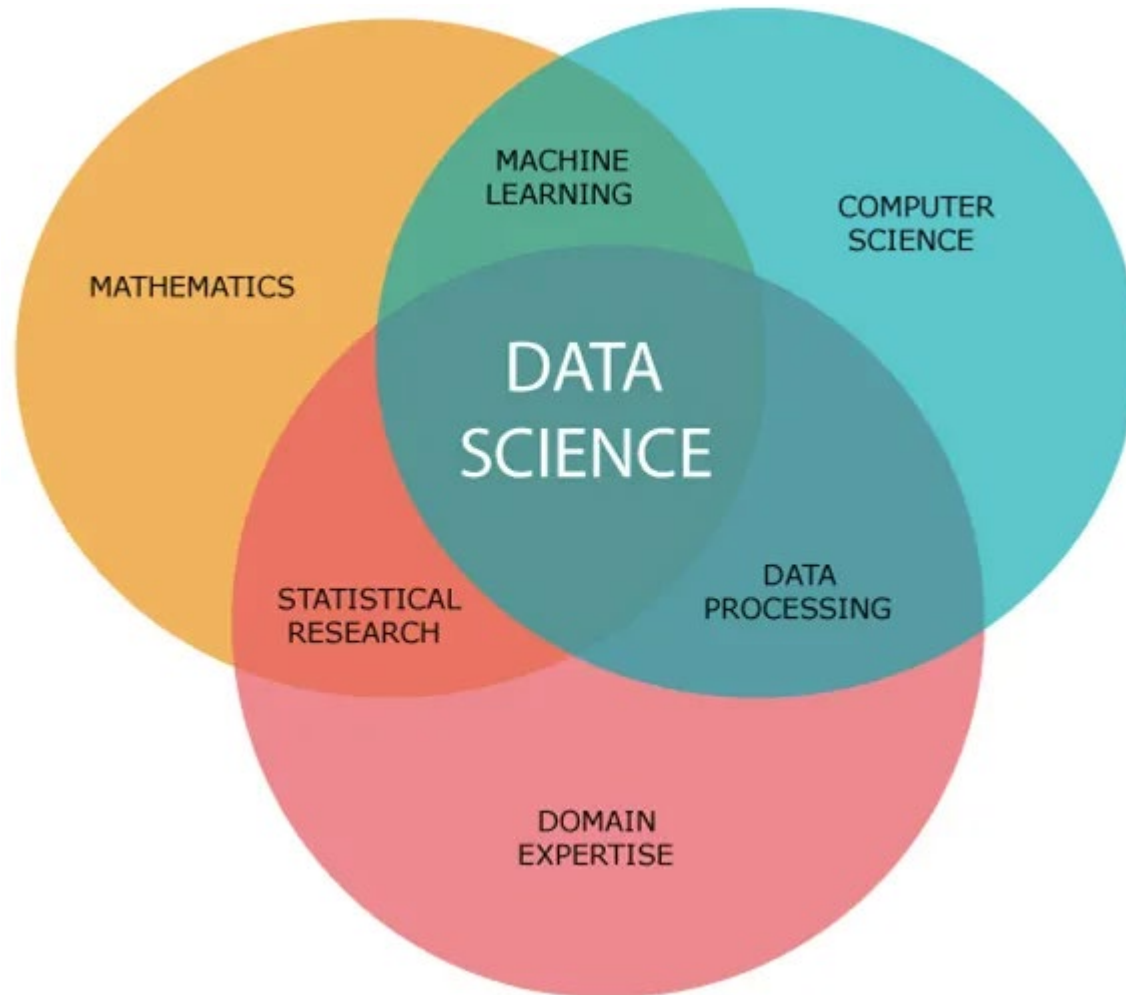


her

Alvaro Pérez Niño - Instructor

¿Qué es la Ciencia de Datos?

Ciencia de Datos



La ciencia de datos es un campo interdisciplinario que utiliza métodos, procesos, algoritmos y sistemas científicos para extraer conocimiento y conocimientos de los datos en diversas formas, tanto estructurados como no estructurados.

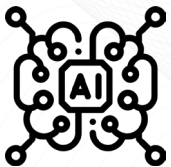
Ciencia de Datos - Elementos clave



Estadística: Uso de técnicas estadísticas para analizar datos y obtener conclusiones.



Análisis de datos: Proceso de inspeccionar, limpiar y modelar datos con el objetivo de descubrir información útil.



Aprendizaje automático: Rama de la IA que permite a las computadoras aprender de los datos y mejorar sus resultados con el tiempo.



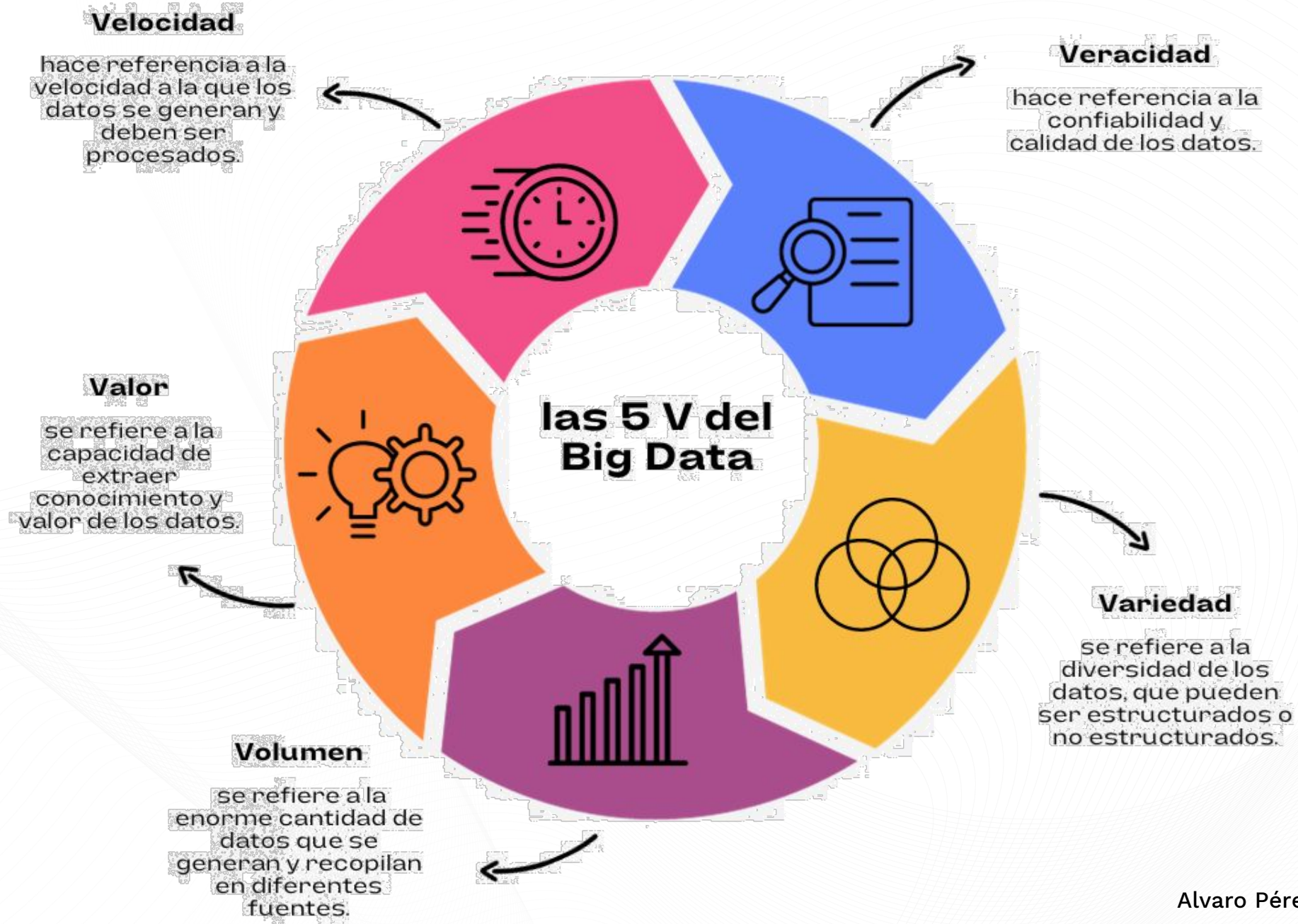
Visualización de datos: Representación gráfica de datos para facilitar su comprensión.

¿Qué es Big Data?

Big Data

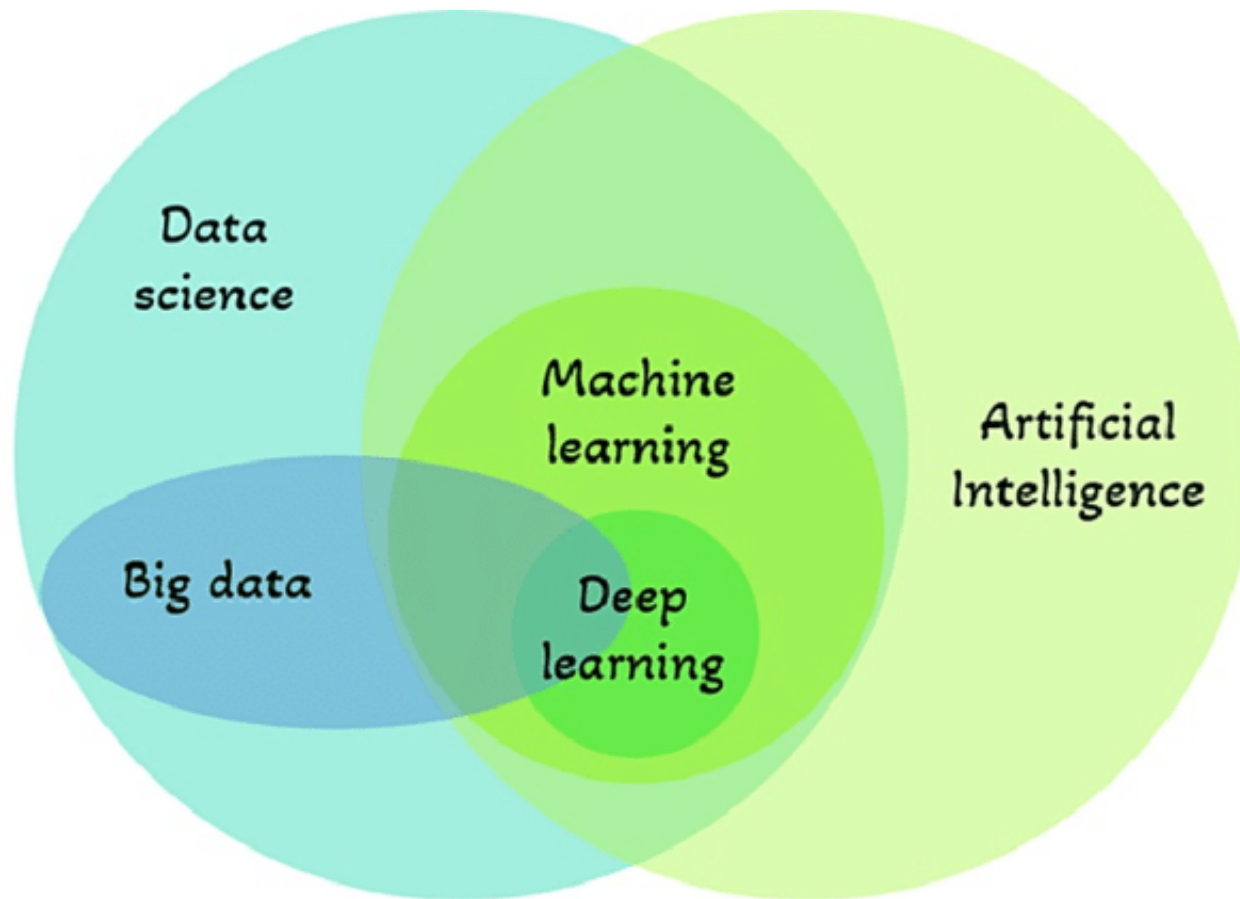
Big Data se refiere a conjuntos de datos que son tan grandes o complejos que las aplicaciones tradicionales de procesamiento de datos no son adecuadas para manejarlos.





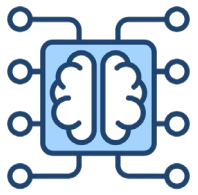
¿Qué es la Inteligencia Artificial?

Inteligencia Artificial



La inteligencia artificial (IA) es la simulación de procesos de inteligencia humana por parte de máquinas, especialmente sistemas informáticos.

Inteligencia Artificial - Áreas clave



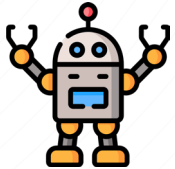
Aprendizaje automático (Machine Learning): Capacita a los sistemas para aprender de los datos y hacer predicciones o decisiones.



Procesamiento del lenguaje natural (NLP): Permite a las máquinas entender y procesar el lenguaje humano.



Visión por computadora: Permite a las máquinas interpretar y tomar decisiones basadas en imágenes visuales (imágenes – video).

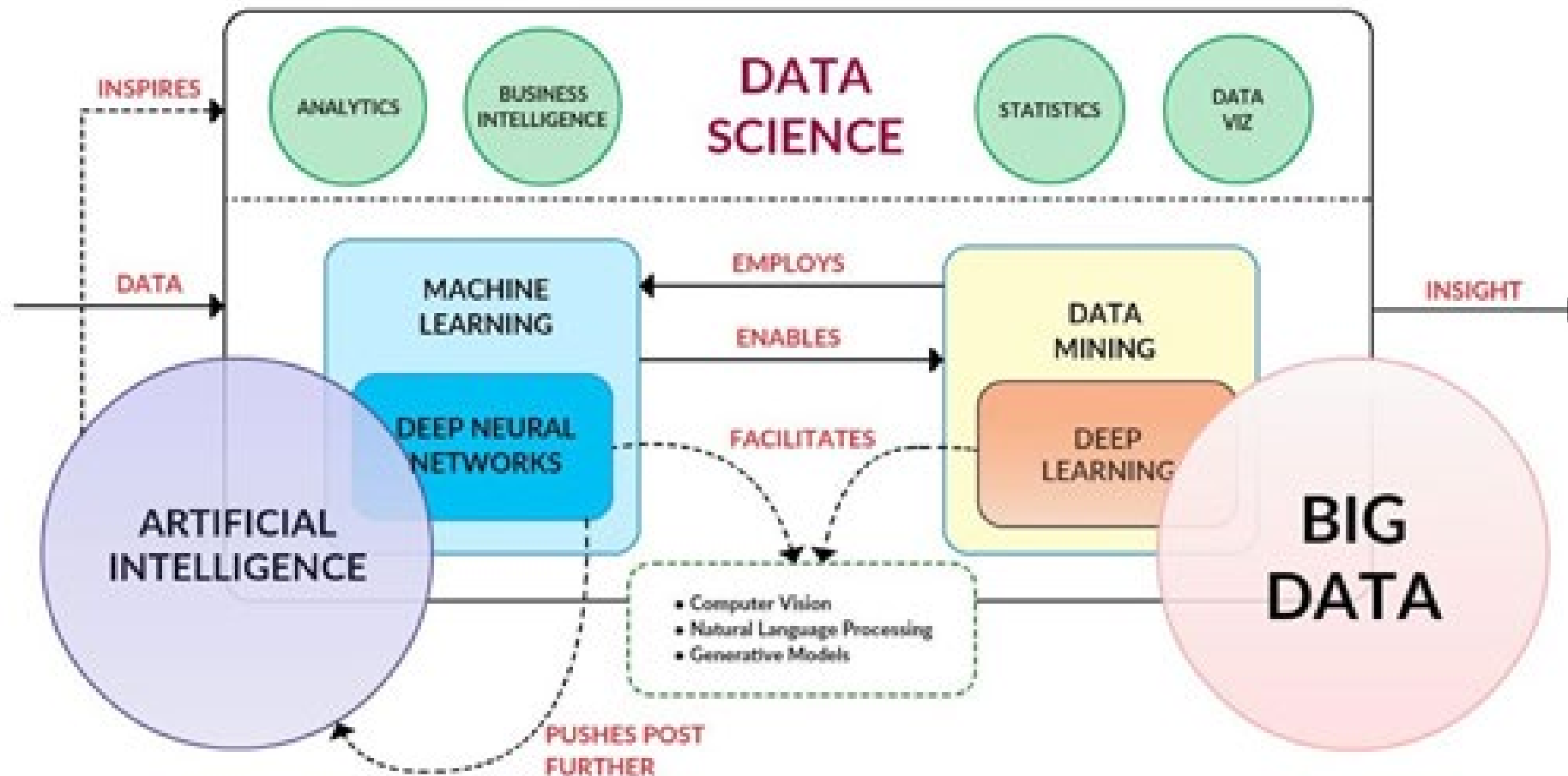


Robótica: Diseño y construcción de robots que pueden realizar tareas humanas.

Relación entre Ciencia de Datos, Big Data e Inteligencia Artificial

- **Ciencia de Datos:** Utiliza Big Data y técnicas de IA para extraer conocimientos.
- **Big Data:** Proporciona los datos necesarios para los análisis y modelos de IA.
- **Inteligencia Artificial:** Se nutre de Big Data y es una herramienta clave en la ciencia de datos.

Relación entre Ciencia de Datos, Big Data e Inteligencia Artificial

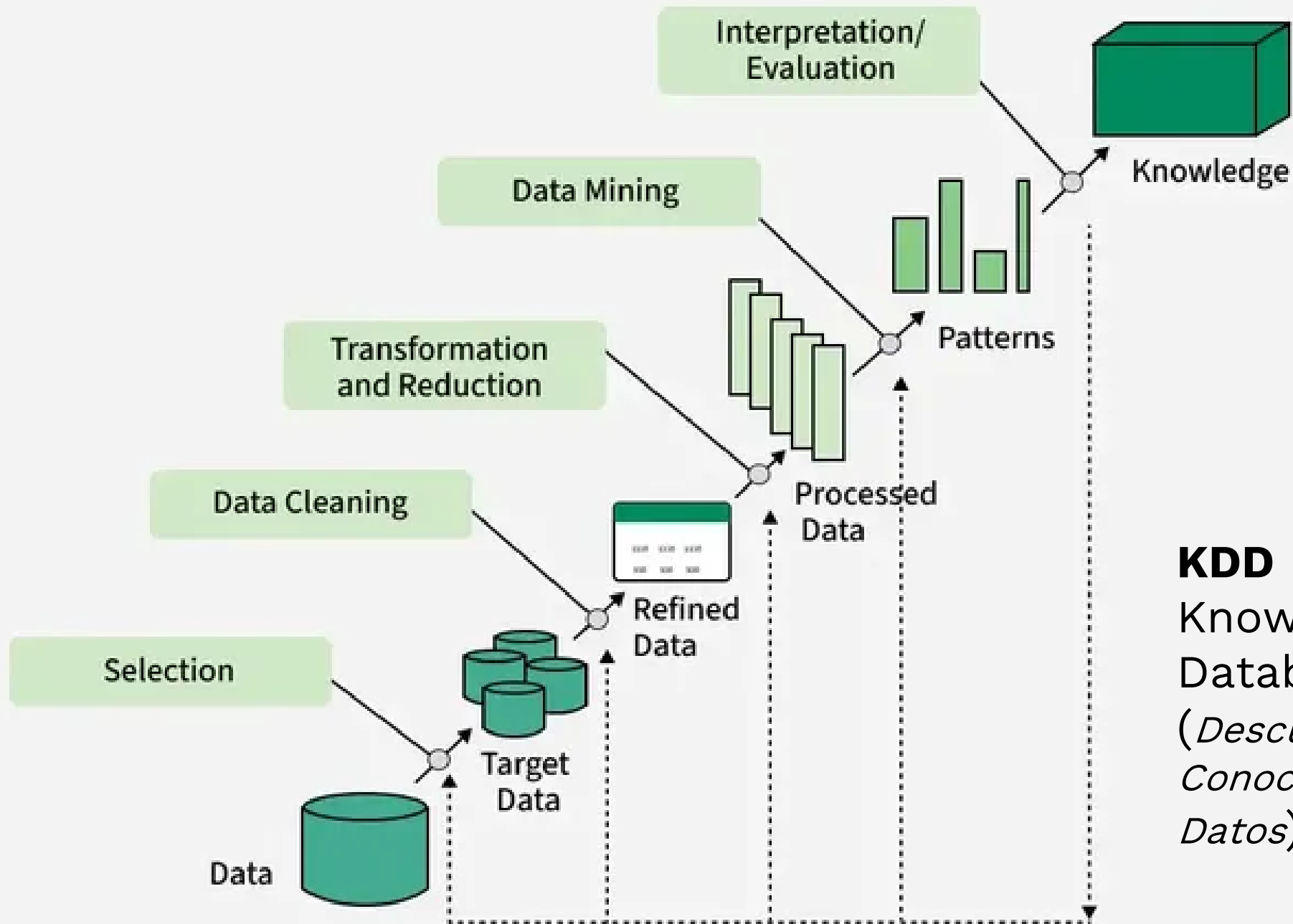


Comparativo



Concepto	Ciencia de Datos	Big Data	Inteligencia Artificial
Enfoque	Analiza datos para obtener conocimiento útil.	Gestiona y procesa grandes volúmenes de datos.	Crea sistemas que aprenden y toman decisiones.
Objetivo	Apoyar la toma de decisiones con base en datos.	Tratar datos masivos con alta velocidad y variedad.	Simular la inteligencia y el aprendizaje humano.
Herramientas	Python, R, Power BI, Tableau.	Hadoop, Spark, NoSQL.	TensorFlow, PyTorch, Scikit-learn.
Aplicaciones	Predicción de ventas, análisis de clientes.	Redes sociales, IoT, monitoreo en tiempo real.	Chatbots, reconocimiento facial, autos autónomos.

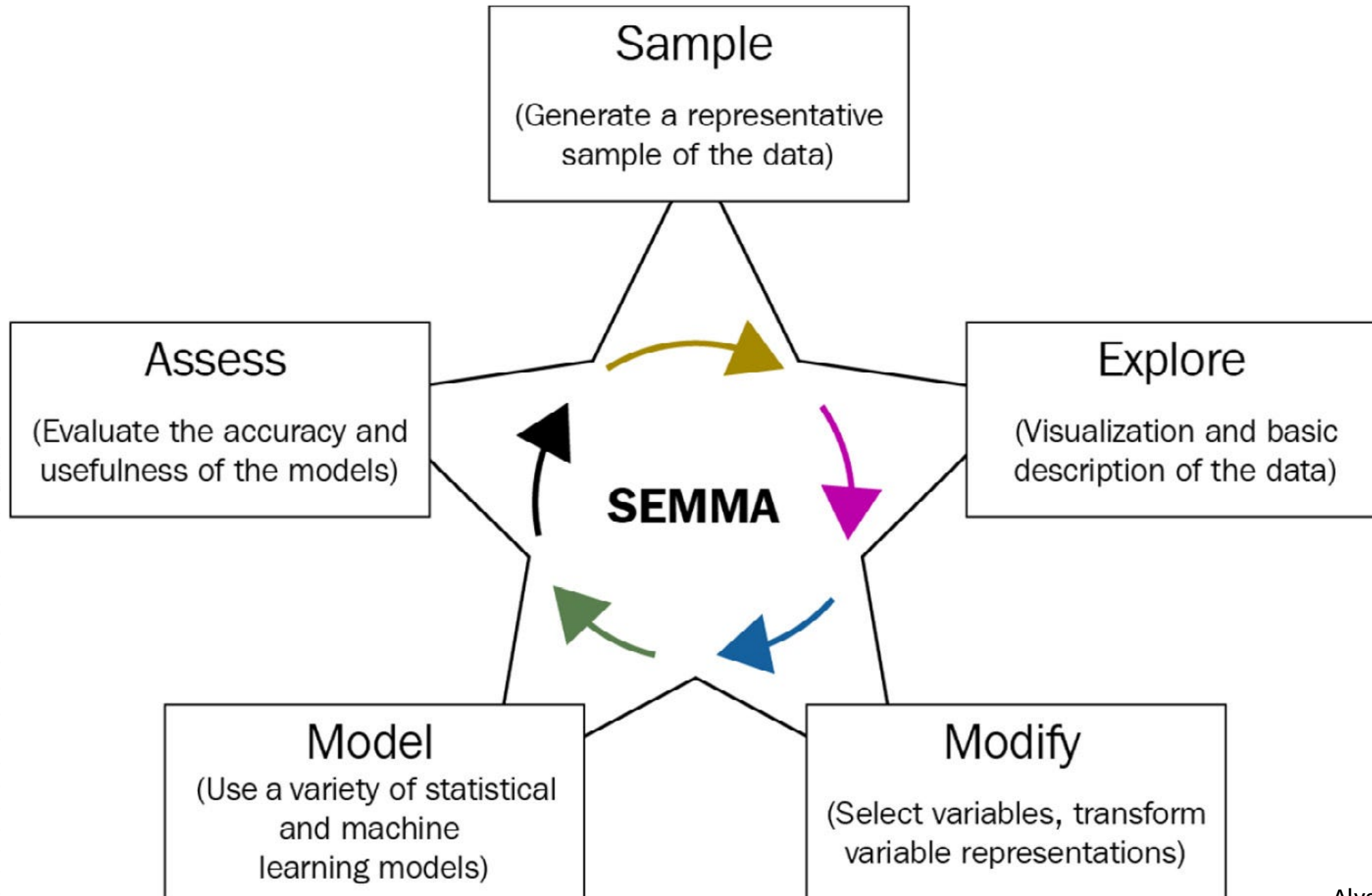
Metodología de Desarrollo

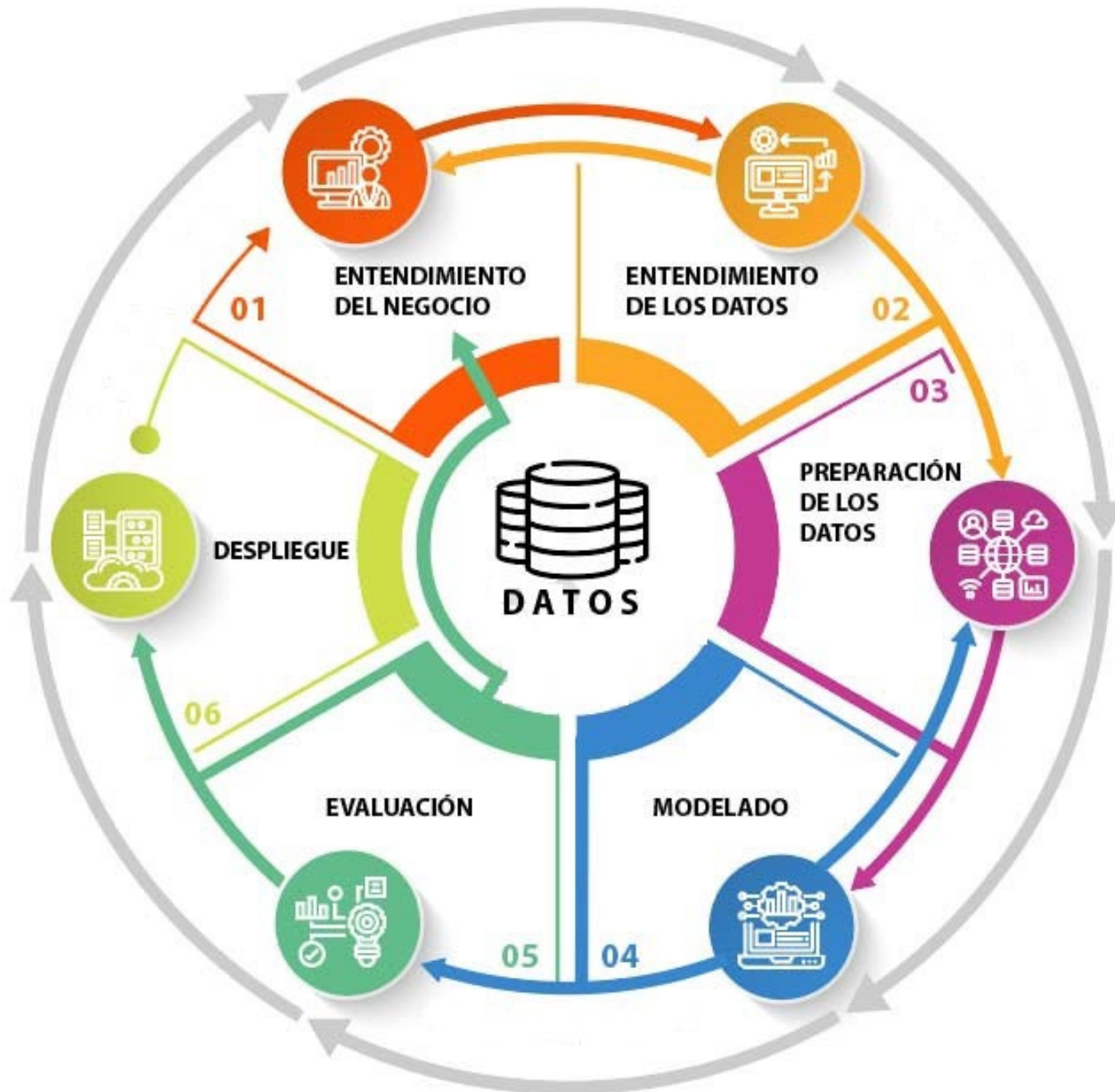


KDD

Knowledge Discovery in Database

(Descubrimiento de Conocimiento en Bases de Datos)





CRISP-DM

*(Cross-Industry
Standard
Process for Data
Mining)*

Perfiles y roles

Rol	Descripción general	Funciones principales	Habilidades requeridas	Herramientas comunes
Científico/a de Datos (Data Scientist)	Profesional que combina programación, estadística y conocimiento del negocio para generar modelos predictivos y tomar decisiones basadas en datos.	<ul style="list-style-type: none"> • Diseñar modelos de Machine Learning. • Analizar grandes volúmenes de datos. • Comunicar resultados mediante reportes y visualizaciones. 	<ul style="list-style-type: none"> • Programación en Python o R. • Estadística y modelado matemático. • Pensamiento analítico. 	Python, R, scikit-learn, TensorFlow, Power BI, Jupyter Notebook.
Analista de Datos (Data Analyst)	Se enfoca en recopilar, limpiar, procesar y visualizar datos para apoyar decisiones estratégicas.	<ul style="list-style-type: none"> • Limpieza y análisis exploratorio de datos. • Elaborar reportes y dashboards. • Identificar patrones o tendencias. 	<ul style="list-style-type: none"> • Excel avanzado. • SQL y Python básico. • Visualización de datos. 	Excel, Power BI, Tableau, SQL, pandas, matplotlib.

Rol	Descripción general	Funciones principales	Habilidades requeridas	Herramientas comunes
Ingeniero/a de Datos (Data Engineer)	Desarrolla la infraestructura que permite almacenar, procesar y transportar grandes volúmenes de datos.	<ul style="list-style-type: none"> • Diseñar y mantener bases de datos y pipelines de datos. • Integrar fuentes de información. • Asegurar calidad y disponibilidad de los datos. 	<ul style="list-style-type: none"> • Bases de datos SQL/NoSQL. • Programación y automatización. • Cloud computing. 	Hadoop, Spark, Airflow, PostgreSQL, AWS, Azure.
Arquitecto/a de Datos (Data Architect)	Define la estructura y políticas del ecosistema de datos de una organización.	<ul style="list-style-type: none"> • Diseñar arquitecturas de datos escalables. • Definir estándares de seguridad y gobernanza. • Coordinar la integración entre sistemas. 	<ul style="list-style-type: none"> • Modelado de bases de datos. • Seguridad y gestión de datos. • Diseño de soluciones cloud. 	Oracle, AWS, Azure, Google Cloud, ER/Studio.

Rol	Descripción general	Funciones principales	Habilidades requeridas	Herramientas comunes
Ingeniero/a de Machine Learning (ML Engineer)	Implementa, optimiza y despliega modelos de aprendizaje automático en producción.	<ul style="list-style-type: none"> • Entrenar modelos a gran escala. • Automatizar pipelines de IA. • Monitorear desempeño del modelo. 	<ul style="list-style-type: none"> • Programación avanzada. • Deep Learning y optimización. • DevOps e implementación. 	TensorFlow, PyTorch, Docker, MLflow, FastAPI.
Analista de Negocio (Business Analyst)	Traductor entre el negocio y el equipo técnico; interpreta los resultados analíticos para proponer mejoras estratégicas.	<ul style="list-style-type: none"> • Comprender necesidades del negocio. • Formular indicadores y métricas. • Presentar hallazgos a directivos. 	<ul style="list-style-type: none"> • Comunicación efectiva. • Análisis financiero o de gestión. • Visualización de datos. 	Power BI, Excel, Tableau, SQL.
Administrador/a de Bases de Datos (DBA)	Gestiona, optimiza y asegura el funcionamiento de las bases de datos de la organización.	<ul style="list-style-type: none"> • Instalar, configurar y mantener bases de datos. • Controlar acceso, rendimiento y seguridad. • Realizar copias de seguridad y recuperación. 	<ul style="list-style-type: none"> • Conocimientos en SQL y optimización. • Administración de sistemas. • Seguridad de datos. 	MySQL, PostgreSQL, Oracle, MongoDB.



GRACIAS

Presentó: Alvaro Pérez Niño
Instructor Técnico

Correo: aperezn@misena.edu.co

<http://centrodeserviciosygestionempresarial.blogspot.com/>

Línea de atención al ciudadano: 01 8000 910270

Línea de atención al empresario: 01 8000 910682



@SENAComunica

www.sena.edu.co