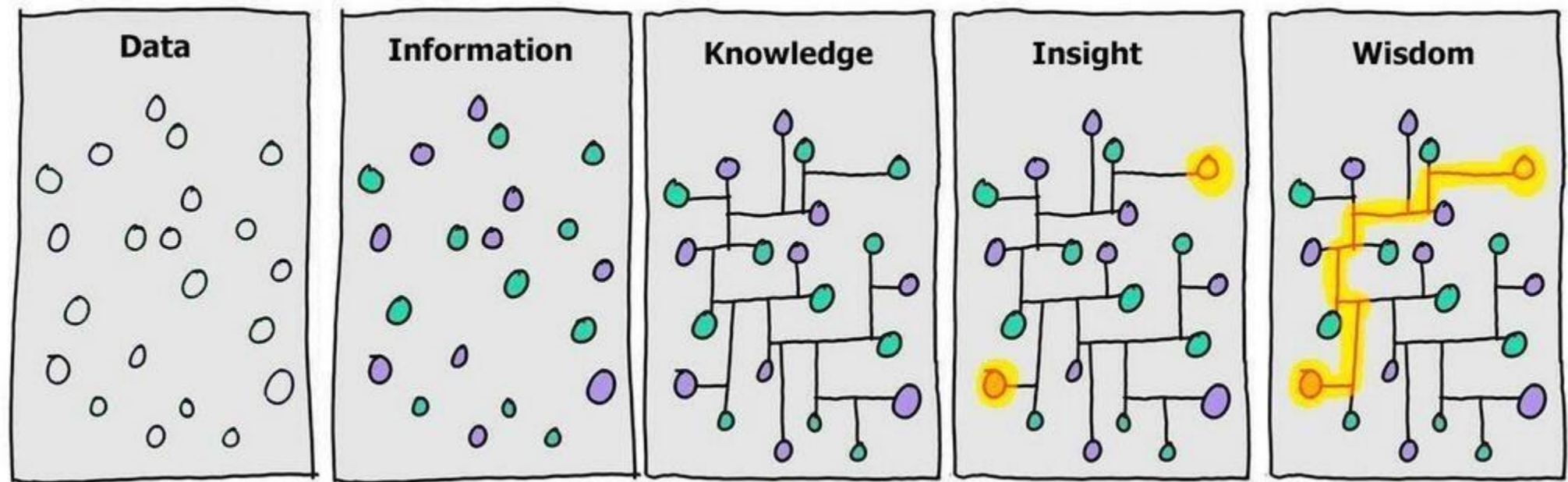


Analítica de datos

Por: Erick Merino



Tipos de análisis

Tipos

- Descriptiva: ¿Qué ocurre?
- Diagnóstica: ¿Por qué X ocurrió?
- Predictiva: ¿Qué ocurrirá?
- Prescriptiva: ¿Qué se debería hacer?

Tipos

- Descriptiva: ¿Qué ocurre?
 - EDA: Análisis exploratorio de datos
 - BI: Business Intelligence
- Diagnóstica: ¿Por qué X ocurrió?
 - Data minning
- Predictiva: ¿Qué ocurrirá?
 - Machine learning
- Prescriptiva: ¿Qué se debería hacer?
 - Optimización
 - Enfoque basado en reglas
 - Aprendizaje por reforzamiento

Análisis Exploratorio de Datos (EDA)

- Primer acercamiento al análisis de los datos
- Toma de muestras estadísticas y de calidad de datos
 - Histograma, cuartiles, percentiles
 - Nulos, ceros, duplicados
- Normalmente se crean gráficos y se toman métricas por cada entidad o conjunto de datos
- Se analizan relaciones entre los datos: correlaciones, otros
- Si son muchos datos, se puede hacer un perfilamiento con una muestra
- Existen herramientas automatizadas: ejemplo pandas profiling, sweetviz

Inteligencia de Negocios (BI)

- Herramientas para crear gráficos, Dashboards y “PPTs” con gráficos con datos
- Suelen usar una estructura de datos de Cubos multidimensionales, pero se pueden usar con datos tabulares y la herramienta construye internamente los cubos



Minería de datos

- Herramientas para encontrar patrones estructurales en los datos
- Herramientas de Clustering o agrupamiento
- Búsqueda de patrones
- Patrones de secuencia
- Otros patrones dentro de los datos

Machine Learning

- Técnicas que se basan en tener un modelo matemático con “huecos” o espacios libres (variables o parámetros), y a través de un proceso de optimización y probabilidad se llenan.
- Hay 2 tipos:
 - Supervisado: El objetivo es predecir una o más categorías o números, ejemplo: Predecir un precio o nota
 - No Supervisado: El objetivo es predecir la función matemática que genera a los datos, es decir un modelo matemático que "explique" los datos, ejemplo: Clustering o Agrupamiento, o las IAs generativas

Póngale play

Deep Learning

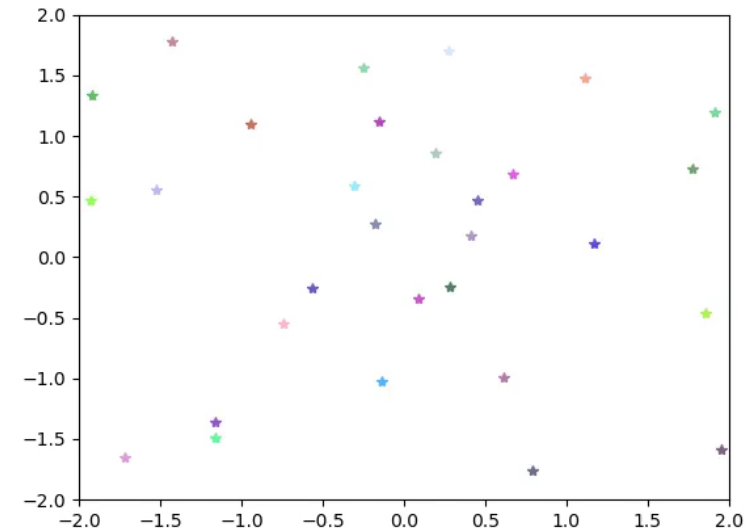
- Un tipo de Machine Learning que utiliza modelos basados en "Redes Neuronales", los cuales son nodos conectados de diferentes formas entre sí. Cada nodo representa una operación matemática como una suma de los valores de entrada multiplicados por un peso, y tiene una salida.
- Según la estructura y forma de las conexiones se dividen en diferentes tipos (convolucional, recurrente, etc)



Optimización

- Algoritmos que dado una función de costo, intentan obtener la combinación de variables óptima tal que el costo sea el menor/mayor
- Pueden ser analíticas, e iterativas, basado en matemáticas o en sistemas biológicos
- Los algoritmos más famosos son:
 - Gradiente Descendiente
 - Fuerza bruta
 - Colonia de hormigas
 - Enjambre de partículas

Póngale play



Aprendizaje por reforzamiento

- Técnica que se nutre del machine learning / Deep learning, y de optimización
- El objetivo es que un “actor” ejecute la “acción” más óptima, dado un “escenario”.
- Se usa en robótica, para hacer que los robots actúen
- Se usa también en autos autónomos
- Algoritmos famosos: Q-Learning

Tipos de redes de Deep Learning

- Las redes neuronales se clasifican según su esquema de conexión entre los nodos. A partir de esto se crean topologías y “Arquitecturas”
- DNN: Red neuronal dirigida/densa, la más básica, todos los nodos de una capa se conectan a todos los de la siguiente
- CNN: Convolutacional, los nodos se conectan imitando la visión local (cómo ve el ojo), a través de capas y filtros convolucionales
- RNN: Recurrentes, tienen nodos que van “al pasado” guardando memoria de los datos anteriores, sirven para series de tiempo

Arquitecturas más famosas en Deep Learning

- Transformers: Red que guarda un peso de atención entre cada dato de una serie de tiempo. Se utilizan para tareas de “Secuencia a secuencia” en generación y traducción de texto, generación de música, etc.
- U-net: Red convolucional con un decoder y un encoder, para trabajos de segmentación de imágenes o generación de imágenes (Las redes para difusión suelen usar u-net).
- Autoencoder: La entrada y la salida es idéntica, se utilizan para comprimir información y descartar datos ruidosos / outliers

Gracias