

APLICACIONES DE MACHINE LEARNING Y CIENCIAS DE DATOS

1. Bienvenida

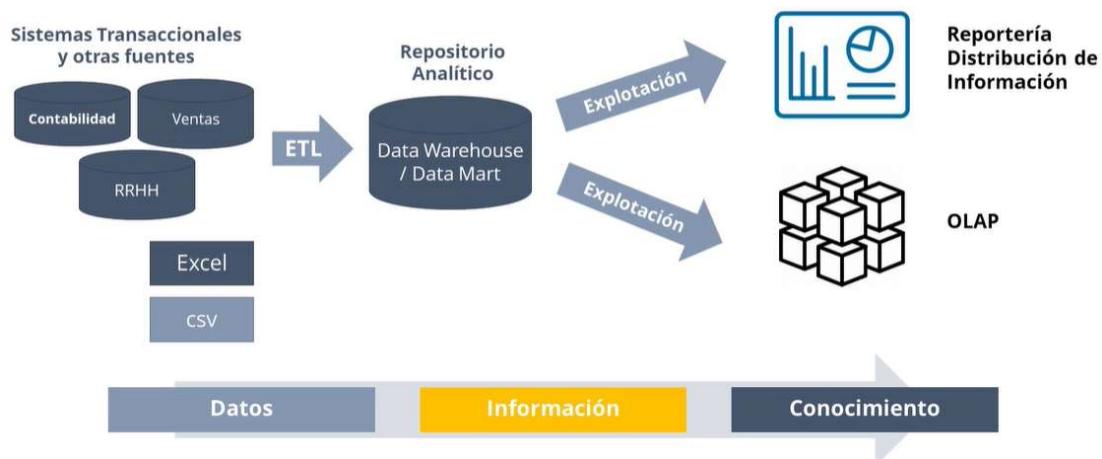
2. Concepto de inteligencia de negocios y minería de datos

a. Inteligencia de negocios:

¿Qué es?

Comprende las estrategias y tecnologías utilizadas por las empresas para el **análisis de datos de negocios**. Su objetivo es permitir la fácil interpretación de estos datos.

- i. La parte de análisis de datos de negocio quiere hacer una contraposición del dato transaccional, de dato operacional que corresponde a una traza o rastro que, en un segundo momento, analítico, se toman, se transforma para extraer información de los datos.
- ii. Principales componentes de una arquitectura de una Inteligencia de Negocio (BI) son:



iii. Sistemas Transaccionales y otras Fuentes:

1. Proveen de datos de la operación del negocio.
2. Son los usuarios sistemas transaccionales de negocio.
3. BI no sugiere su cambio, sino tomarlos y pasarlos por un proceso llamado ELT (Extracción, Transformación y Carga). Esto permitirá seleccionar las variables de mayor interés, generar las transformaciones que permiten homogeneizar formatos, criterios, formas de cálculo, formulas, ... etc., que permitir hacer una carga sobre un nuevo sistema que será analítico.
4. El repositorio analítico será la fuente y la base de toda la inteligencia corporativa.
5. Distinto al mundo transaccionar donde la inmediatas y la continuidad son los factores claves. En el mundo analítica no se busca estas métricas, sino construir sistemas que permite una rápida extracción de información.
6. A través de la explotación, es decir, del buen uso de este repositorio se obtiene la repostería distribución de información.
7. También se obtienen los sistemas OLAP (Análisis Online de Procesos) que son tipos de BD que permiten tener precálculos de algunas de las variables más críticas del análisis del negocio.

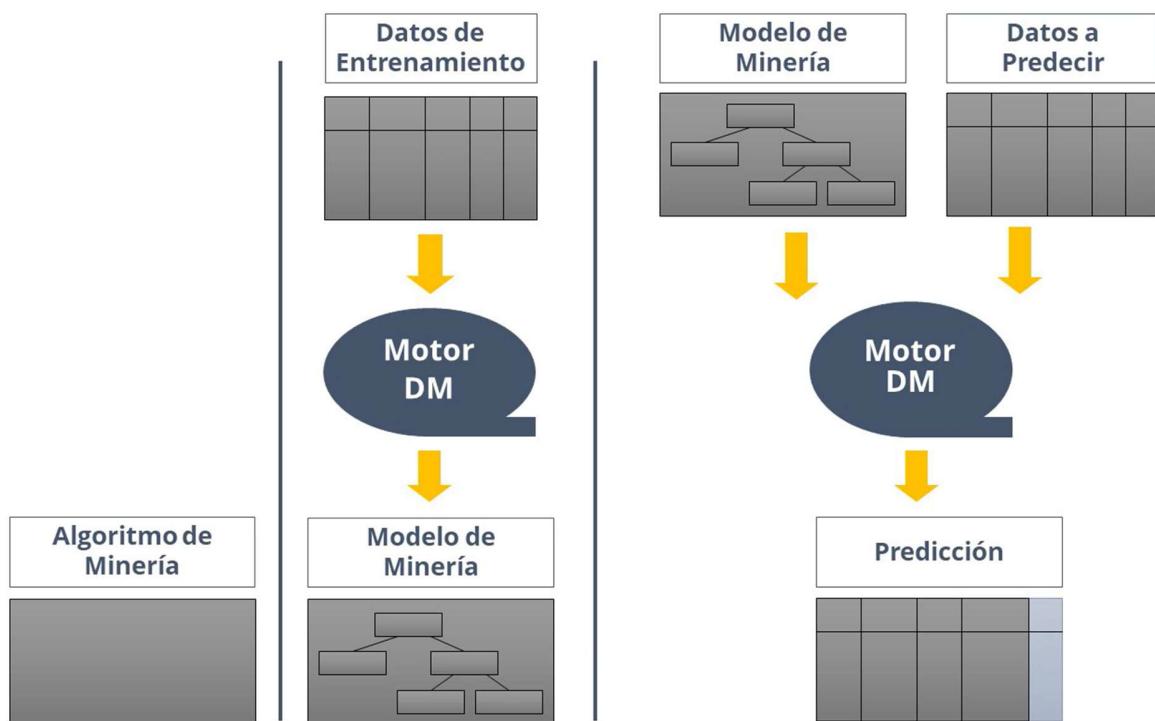
b. Minería de datos/Ciencia de Datos (DM + DS):

¿Qué es?

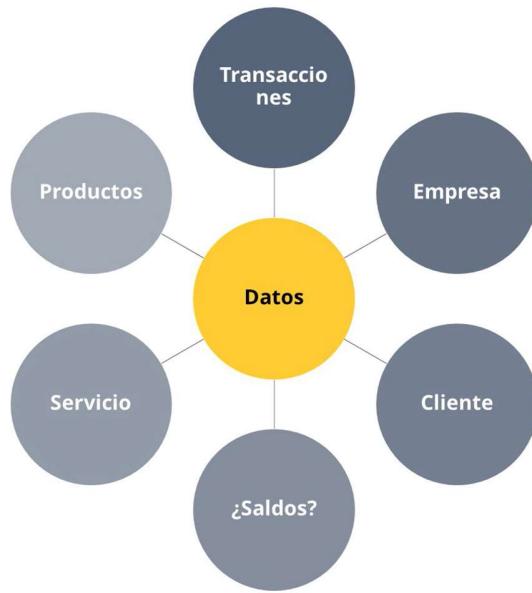
Es el descubrimiento de conocimiento en un gran conjunto de datos. El conocimiento que se obtiene viene dado en forma de **correlaciones, patrones y tendencias**. Utiliza técnicas de reconocimiento de patrones, matemáticas y estadísticas.



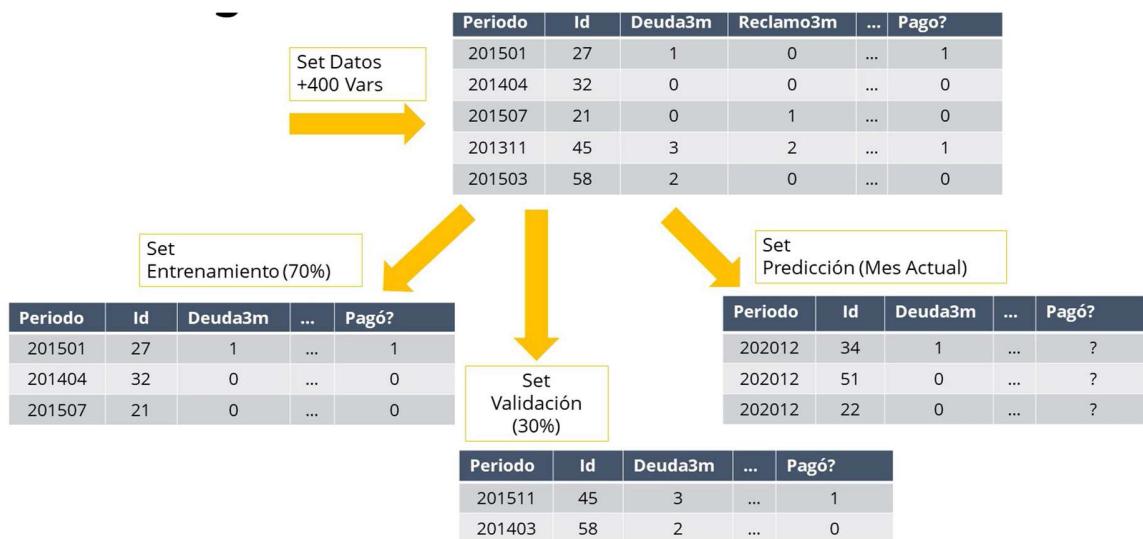
i. Metodología:



1. La metodología comienza con la recolección de datos:



2. Luego de los datos, se crea un tablón sobre el cual se puede ejecutar el resto de la metodología:



- Inicia con 2 variables, un periodo de tiempo del que hablamos y un identificador del cliente. El cliente y el periodo pueden aparecer muchas veces en el DataSet.
- Luego, los descriptores, ejemplo, edad, sexo, deudas.
- Luego, la última columna, un número que indique si ocurrió o no.

3. Luego, el dataset se divide en 3 grupos:

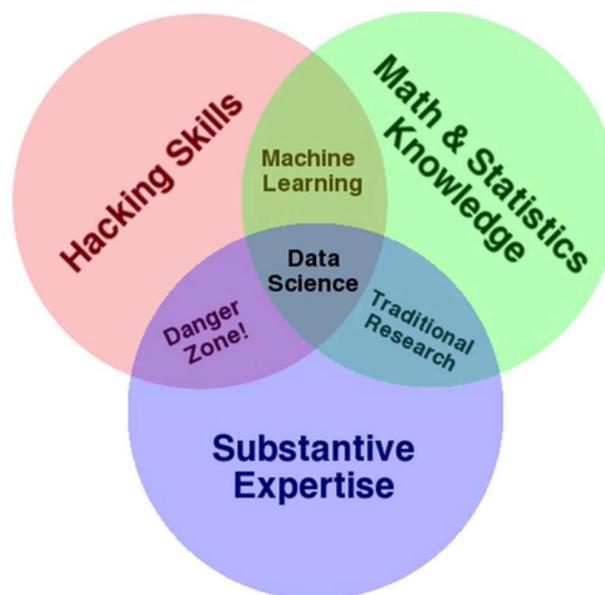
- a. Set de entrenamiento (70%): búsqueda de patrones.
- b. Set de validación (30%): validar el entrenamiento.
- c. Set predicción (mes actual).

3. Conceptos de Ciencias de Datos y Aprendizaje de Máquina

a. Ciencias de Datos (DS):

Campo interdisciplinario, involucra métodos, procesos y sistemas para extraer conocimiento a partir de datos en sus diferentes formas, estructurados o no estructurados (Big Data)

Relacionado con transformación de Datos en Insights



- i. Según el diagrama de Conway, un Data Science debe tener conocimiento en:
1. Matemáticas y conocimientos en estadísticas.
 2. Habilidades en programación, infraestructura y tecnología. Pero no estudios formales, si no, son truco, técnicas y elementos que se aprovecha para completar un análisis da datos.
 3. Conocimientos sustantivos del dominio, es decir, conocer el negocio.

b. Aprendizaje de máquina (ML):

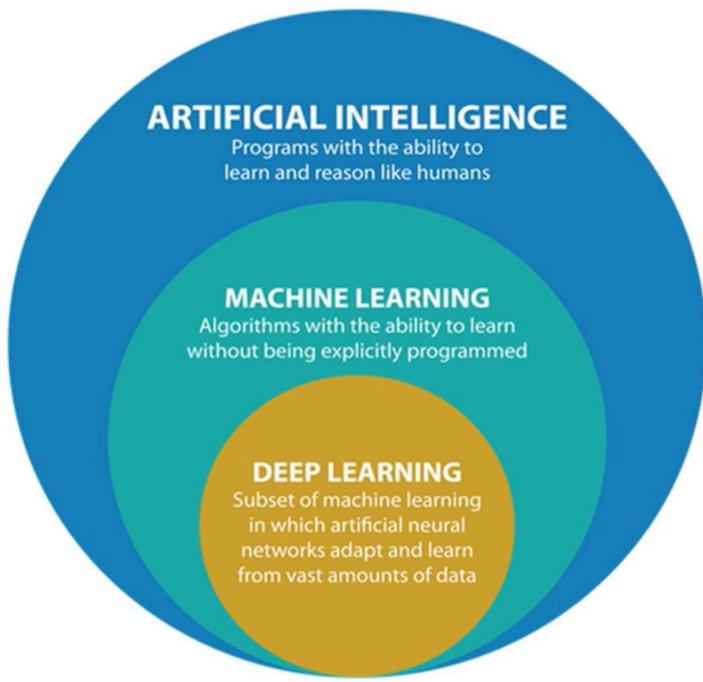
Es la rama de la inteligencia artificial, cuyo objetivo es desarrollar técnicas que permitan que las computadoras aprendan

Algoritmos y heurísticas para convertir muestras de datos en programas, sin tener que escribir los últimos explícitamente

- i. Para apoyar la definición:

Inteligencia Artificial

Concepto aplicado a una máquina al imitar las funciones «cognitivas» **sin ser explícitamente programado** para hacerlo.



1. Dentro de la IA, se encuentra una subrama llamada **Machine Learning (Aprendizaje de Máquina)**, que utiliza particularmente métodos estadísticos de aprendizaje que aprenden desde la historia de datos que este recolectando.
2. Dentro de ML se encuentra una subrama llamada **Deep Learning (Aprendizaje Profundo)**, utiliza una técnica estadística llamada redes neuronales profundas.

ii. Como se relacionan la DS con la IA:

Ciencia de Datos
<p>Campo interdisciplinario, involucra métodos, procesos y sistemas para extraer conocimiento a partir de datos en sus diferentes formas, estructurados o no estructurados (Big Data)</p>

1. La ciencia de dato va a intentar la extracción de conocimiento a partir de los datos, para esto, puede utilizar herramientas de IA, ML, y DL.
2. El fin de la DS es buscar a partir de los datos un patrón oculto que genere impacto en el negocio.
3. En DS si bien se puede buscar un fin predictivo, es muy importante el patrón, en tratar de entender cual es esa regla, ese elemento común que se repite y que permite eventualmente mejorar el proceso.
4. En ML el foco está en la búsqueda de la predicción (ajuste) y para eso, sacrifica el patrón para mejorar la predicción. Foco más efectivo que permite buscar un mejor predictor que falle muy poco. Para esto se generan modelos oscuros llamados modelos de cajas negras donde no se conoce la estructura, ni la claridad del DS, pero se gana mucho en la reducción del error de la predicción.

4. El auge actual del Aprendizaje de Máquina

a. Inteligencia artificial:

Una máquina imita las funciones «cognitivas» como "aprender" o "resolver problemas" **sin ser explícitamente programado para hacerlo.**

- i. La IA se hace crítica en varios problemas llamados **inefables**, que son difíciles de resolver usando programación procedural.



ii. ¿Cuándo utilizar IA?

1. Cuando un proceso o decisión es muy repetitivo.
2. Cuando una solución carece de una solución explícita.
3. Cuando existen muchos datos disponibles para entrenar.

5. Aplicaciones sobre transacciones estructuradas

a. Sistemas transaccionales:

- i. Sistemas que abarcan toda la empresa u organización.



ii. Los procesos son (ejemplo):

1. Registrar transacciones de un supermercado.
2. Pagar facturas a los proveedores.
3. Pagar sueldo a los funcionarios.
4. Avisar de faltas de stock.

iii. Son el corazón de las organizaciones ya que permiten que los procesos operen.

iv. Estos sistemas están diseñados para:

1. Tener altísima disponibilidad.
2. Contar con redundancia en caso de fallas.
3. Utilizar el mínimo espacio posible (uso de mucha codificación).
4. Registrar todo el detalle posible.
5. Ser alterados por muchas aplicaciones.

v. Bases de datos: permite registrar datos estructurados. Colección de tablas que permite almacenar registro.



Colección única y estructurada de datos



Describe las actividades de una organización o de un área de ella

Si el sistema es altamente estructurado,



Se valida su consistencia con la lógica del negocio

Modelo Entidad - Relación



- Técnica de diseño de BD gráfica.
- Muestra información relativa a los datos y sus relaciones.

Modelo Entidad - Relación



- 1 Fácil de entender
- 2 Buena **herramienta de comunicación**
(entre diseñadores y usuarios no técnicos)
- 3 Luego del **diseño conceptual**, se **lleva a cabo un diseño lógico. (SGBD)**

vi. 3 elementos componen un modelo Entidad – Relación:

1. **Entidades**: representa un objeto o concepto del mundo real. Ej. “Empleado”, “Proyecto”.
2. **Atributos**: representa alguna propiedad de interés que da una descripción más amplia a la entidad. Ej. “Nombre” y “Salario” de una entidad “Empleado”.
3. **Relación**: vínculo entre dos o más entidades. Describe una interacción entre entidades, relación “Trabaja en” entre “Empleado” y “Proyecto”.

vii. Cardinalidad de relaciones:

1. Es el número de instancias o vínculos en que puede participar una entidad.

Razón de cardinalidad para vínculos binarios

Mostrado en diagramas ER por medio de un 1, M, y N en los rombos de las relaciones

Vínculos Binarios

1:1 "Dirige" entre Empleado y Proyecto

1:N "Trabaja en" entre Empleado y Proyecto

M:N La misma que la anterior, pero si los proyectos pudiesen contar con M empleados

- viii. Finalizado el modelo entidad – relación, un diseñador de BD puede generar un **modelo relacional**. Este modelo ya puede ser vertido en un motor de base de datos.

6. Aplicaciones sobre texto

a. **Datos de texto y minería de texto:**

- i. Los datos de texto que se usan como fuentes pueden ser páginas web, libros, correos electrónicos, reseñas de clientes, artículos, entre otros.
- ii. Para analizar estas fuentes de datos escritos se usan la minería de textos.
- iii. La **minería de textos** es una rama de la minería de datos que se refiere al proceso de analizar y derivar información nueva de textos.
- iv. Por medio de la identificación de patrones o correlaciones entre los términos se logra encontrar información que no está explícita dentro del texto.
- v. Ejemplo de analítica de texto son la declaración en texto de los accidentes de trabajo y trayecto. En este tipo de situaciones la identificación de elementos clase rondan entre:

Criterio	Precisión
Unidad anatómica	>90%
Agente externo	>85%
Forma	>90%
Causa	>70%

vi. Otro ejemplo, son los modelos de lenguaje natural (Turing-NLG).

- El resumen que acaba de leer no fue escrito por los investigadores, ni por los periodistas. Fue escrito por el mismo modelo de IA que acabamos de describir.
- Como los modelos anteriores de Aprendizaje Profundo para el lenguaje, se alimenta de prácticamente cualquier texto escrito en Internet.
- Siguiendo la tendencia de que los modelos más grandes llevan a mejores resultados, T-NLG es el modelo más grande publicado hasta la fecha.
- Entre sus posibles usos se encuentran generar palabras, completar frases, comprender documentos, responder a preguntas, correos electrónicos y hacer resúmenes; todo ello de una forma cada vez más fluida y natural.

vii. Algo de contexto:

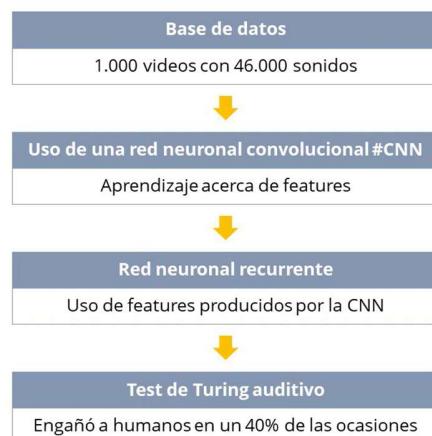
1. +300M bots activos.
2. +90.000.000 de Smart Speakers (Alexa/Google/Cortana/Siri).
3. Según Gartner el 85% de las interacciones por servicio va a ser analizada o tratada por algún agente virtual.
4. +500.000 desarrolladores de temas de análisis de texto.
5. 40% de los milenarios usan por lo regular Forbes.
6. Según Microsoft existen 12 lenguajes para el análisis de texto.

7. Aplicaciones sobre audio

- a. El registro de audio se hace registrando la presencia de distintos pulsos en el tiempo de alta frecuencia de tal manera que es capaz de capturar elementos de distintas frecuencias (alta – agudo, baja – grave).
- b. Una aplicación es la eliminación del ruido (Denoising), separa los sonidos.
- c. Otra aplicación, la predicción de sonido, como a partir de una imagen se puede inferir cual es el sonido que está ocurriendo.

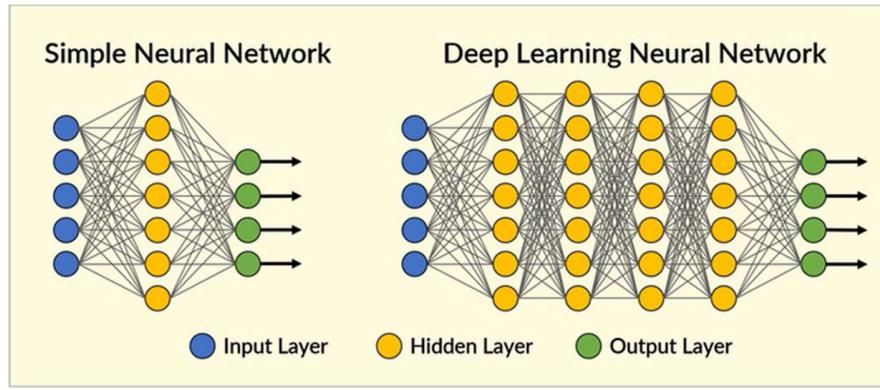
Predicción de sonido

Completar audio



8. Aplicaciones sobre video

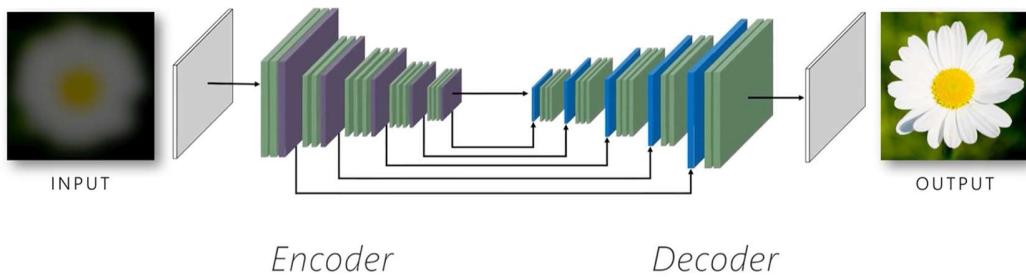
- a. Los videos que corresponden a un conjunto de imágenes almacenadas en el tiempo, se guardar en pixeles, que son puntos que capturas una intensidad de color en un lugar específico dentro de la imagen.
- b. Se descomponen en tres capas (RGB).
- c. Se utilizan las redes neuronales profundas. Estas toman una capa de entrada y va produciendo cálculos en las capas ocultas, así sucesivamente hasta llegar a la capa de salida.



- d. Si se cuenta una DB grande de Input Layer, se podrían generar capas ocultas para identificar la imagen de salida.
- e. Una imagen no solo una hoja de pixeles, son 3 capas de pixeles que representa a los colores primarios.
- f. Una red neuronal de análisis de imágenes se conoce como **ConvNets** (red convolucional).
- g. Alucinaciones (Deep Dreams), se permite a la red neuronal modificar la imagen.

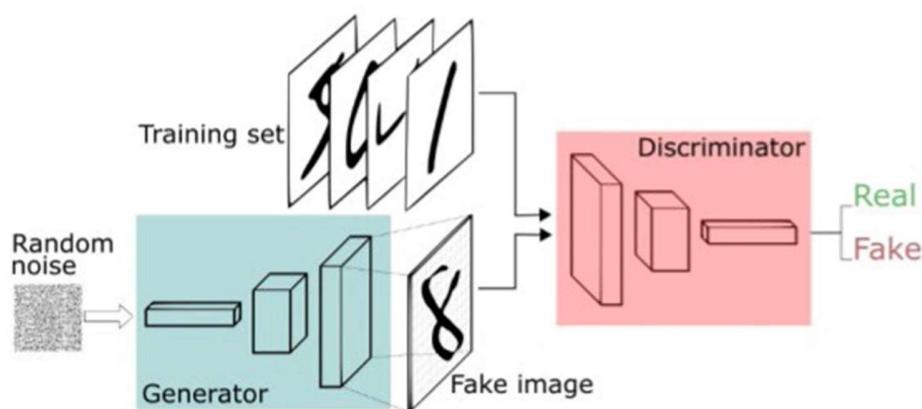


- h. Completar pixeles y frames. Se utiliza topología de redes neuronales convolucionales profundas llamada Encoder Decoder. ¿Cuál es la alucinación qué la red debe hacer de manera que el frame que interpola entre una imagen y otra haga máxima coherencia, no solo con la imagen que está viendo, sino con otros videos e imágenes que el conoce?



i. Completar color (GAN Architecture)

GAN Architecture



i. Se compone de 2 partes:

1. Parte generadora, a la que le dan la capacidad de producir imágenes, de alucinar.
 2. Parte discriminadora, aprende a discriminar si la imagen es generada por la red neuronal y por un conjunto de datos de entrenamiento.
- ii. Esto permite por ejemplo la colorización, donde los datos de entrenamiento corresponden a muchas imágenes a color.
- iii. El lado generador recibe una imagen en B/N y se le permitirá hacer cambios hasta que se logre engañar a la red discriminadora, es decir, esta red entrara imágenes a la red discriminadora hasta que esta red piense es correctamente a color.

Completar color

- 4 No es necesario entregar más información adicional
Foto original en blanco y negro.
- 5 Reparar color
Tecnología utilizada para reparar fragmentos perdidos, reconstrucción 3D, e interpolación de video.
- 6 Imaginar muchas otras aplicaciones positivas de esta tecnología



- j. Animar imágenes.
- k. Video Analytics en tiempo real: utilizando matemáticas, estadísticas e IA, se desarrollan modelos avanzados para “enseñar a un programa a ver y entender el mundo como lo haría una persona”.
- I. Distanciamiento social.

AYUDANTIA 1:

- **Inteligencia de negocios:** Son las estrategias, aplicaciones de datos, productos y tecnología utilizadas en una empresa para el análisis del negocio.



- **Minería de datos:** Es el descubrimiento o conocimiento que se obtiene de un gran conjunto de datos.

- **Ciencia de datos (DS):** Campo que involucra métodos, procesos y sistemas para extraer conocimiento de datos en diferentes formas (Estructura, texto, audio, video).
- **Aprendizaje de máquina (ML):** Es el aprendizaje automatizado de máquinas. Algoritmos para convertir muestras de datos en programas sin tener que codificar.
- **Datos estructurados:**

- Puede guardarse en filas y columnas.
- Abarcan toda la empresa u organización.
- Tienen alta disponibilidad.
- Utilizan el mínimo de espacio posible (mucho código).
- Detallados.
- Ejemplo:
 - Registro de transacciones de un local.
 - Registro de transacciones de un cliente.
 - Cuentas de luz, agua, gas.
 - Aplicación, predicción de ventas.

- **Datos no estructurados – Tipo texto:**

- Son fuentes completas de texto.
- Para analizar este tipo de dato se buscan patrones que permitan extraer información.
- Ejemplo:
 - E-Mail, páginas web.
 - Libro, mensaje de texto.
 - Aplicaciones:
 - Accidentes de trabajo y trayecto.
 - Completar texto en un correo electrónico, WhatsApp.

- **Datos no estructurados – Tipo audio:**

- Distintos pulsos en alta frecuencia.
- Se miden las frecuencias auditivas en un periodo de tiempo.

- Ejemplo:
 - Audio de WhatsApp.
 - Música.
 - Radio.
 - Aplicaciones:
 - Eliminación de ruido.
 - Siri/Alexa.
 - Agregar audio a videos.
- **Datos no estructurados – Tipo video:**
 - Conjunto de imágenes en un tiempo determinado.
 - Se guarda en pixeles.
 - Para analizar este tipo de datos se requieren Redes Neuronales Profundas.
 - Ejemplo:
 - Cámara de un supermercado.
 - Videoclip.
 - Aplicaciones:
 - Agregar color a un video en B/N.
 - Determinar la distancia social en un conjunto de personas.
 - Determinar la edad de las personas que entran a una tienda.

Modulo 3 y 4:

Visualización en ciencias datos:

1. La importancia de la visualización:

- a. Punto 1: Utilizar** el sistema visual humano para extraer la información a partir de los datos.
- b. Punto 2: Proporcionar** descripción de datos complejos.
- c. Punto 3: Identificar** las estructuras y los patrones.

2. Ejemplo: Mapa de Minard 1861.

3. Ejemplo para el miniproyecto, mapeo pandemia, crecimiento.

4. Tipos de visualización:

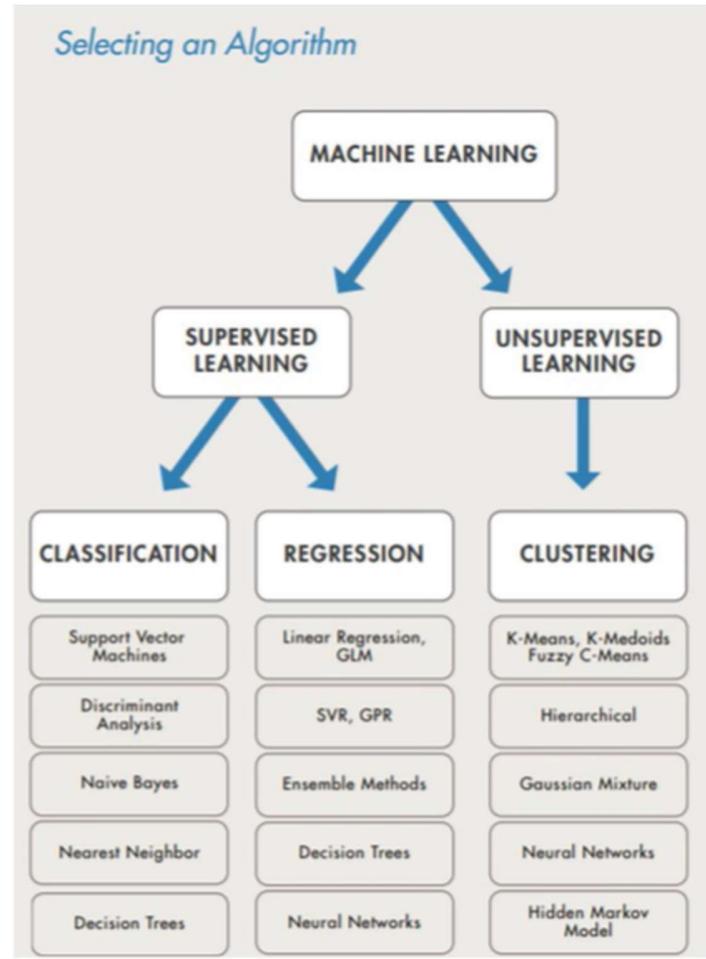
- a. Business Scorecard:** busca ofrecer vista del funcionamiento solo en una pantalla (KPI: Key Performance Index).
- b. Dashboard:** es más táctico, permite, personalizar la vista de instrumentos.
- c. Reportes Estáticos:** permite compartir información de alto detalle.
- d. Reportes Dinámicos:** permite dar capacidad analítica, y permite construir nuevas visualizaciones, gráficos y cuces.
- e. Visualización Avanzada:** permite abarcar ciertos objetivos específicos.

Aprendizaje supervisado:

1. El sistema aprende basándose en datos estructurados o no estructurados, previamente clasificados/etiquetados.

- a. Data:** (datos, etiqueta).
- b. Objetivo:** aprender una función para mapear $x \rightarrow y$
- c. Ejemplo:** clasificación, regresión, Detección de Objetos, etc.

2. Tipos de algoritmos:



a. Algoritmos de clasificación:

- i. Etiqueta un caso. Predice algo, por ejemplo, esto es un perro dentro de muchas fotos.

b. Algoritmos de regresión:

- i. El objetivo es predecir un valor continuo. Por ejemplo, predecir temperatura, renta.

3. Ejemplo, Árbol de decisión:

- a. Diagramas lógicos que representan una serie de reglas de decisión sucesivas.
- b. La forma de estos árboles facilita entendimiento.
- c. Un árbol de decisión partitiona el espacio formado por las variables en un conjunto de hiper-rectángulos.
- d. En cada hiper-rectángulo se ajusta un modelo sencillo, generalmente una constante. $y=c$.
- e. Al hiper-rectángulo se le asigna una regresión línea para poder hacer una predicción.
- f. Pueden utilizarse valores discretos o continuos:

- i. Valores discretos -> árbol de clasificación.
- ii. Valores continuos -> árbol de regresión.

Aprendizaje no supervisado:

1. No hay datos de entrada y no hay clasificación, ni etiquetas.
2. **Modelo** construido utilizando datos como entrada, sin clasificación previa.
3. **Solo datos**, no hay etiquetas.
4. **Objetivo**, aprender alguna estructura oculta subyacente de los datos.
5. **Ejemplo**, algoritmos de agrupación, aprendizaje de características, segmentación de imágenes.
6. Todos los algoritmos caen en los llamados algoritmos de Clustering.
7. Una aplicación de este tipo de algoritmos es la **Segmentación**:
 - a. El proceso por el cual una empresa partitiona el mercado en sub-mercados de tal forma que la respuesta a las variables de marketing varía sustancialmente entre distintos segmentos, pero muy poco dentro del segmento.
 - b. En general los consumidores son muy numerosos, dispersos y necesidades muy heterogéneas.

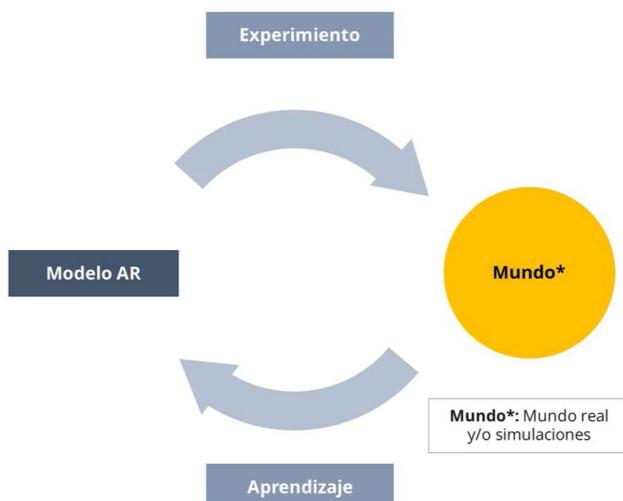
Aprendizaje reforzado:

1. Permite entrenar en escenarios donde no se tengan una recolección histórica de datos.
2. **Definición**, es un conjunto de técnicas basados en psicología del comportamiento, que buscan determinar que acciones debe elegir un agente en un entorno para maximizar una recompensa.
3. **Data**, no se requieren datos históricos.
4. **Objetivo**, aprender a tomar medidas para maximizar la recompensa.
5. **Ejemplo**, control automático, juego Atarí, Sistemas autónomos, Alpha Zero.
6. Tipos de aprendizajes:

Tipo de Aprendizaje	Datos	Etiquetas
Supervisado	SI	SI
No supervisado	SI	NO
Reforzado	NO	NO

Aprendizaje reforzado

Definición
Conjunto de técnicas que permiten a la IA experimentar y aprender.
Ciclo de aprendizaje
Ocurre a velocidad digital y en tiempo real*.
Experimentar en el mundo*
Permite descubrir nuevos patrones.

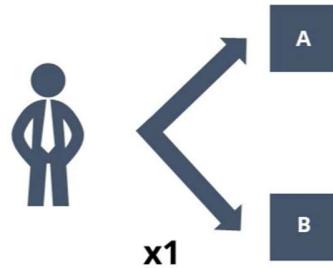


7. Explotación vs Exploración:



8. A/B Testing vs Aprendizaje Reforzado:

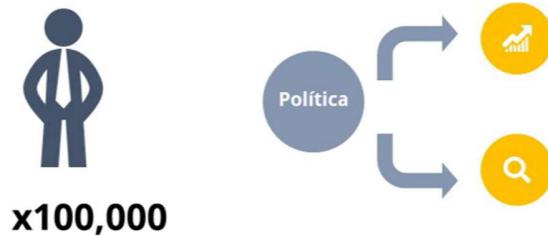
- a. A/B Testing, presenta un desafío a un diseño.



A/B Testing

- 1 Diseña el experimento adecuado.
- 2 Prueba en línea una vez, obtener datos.
- 3 Empieza de nuevo.

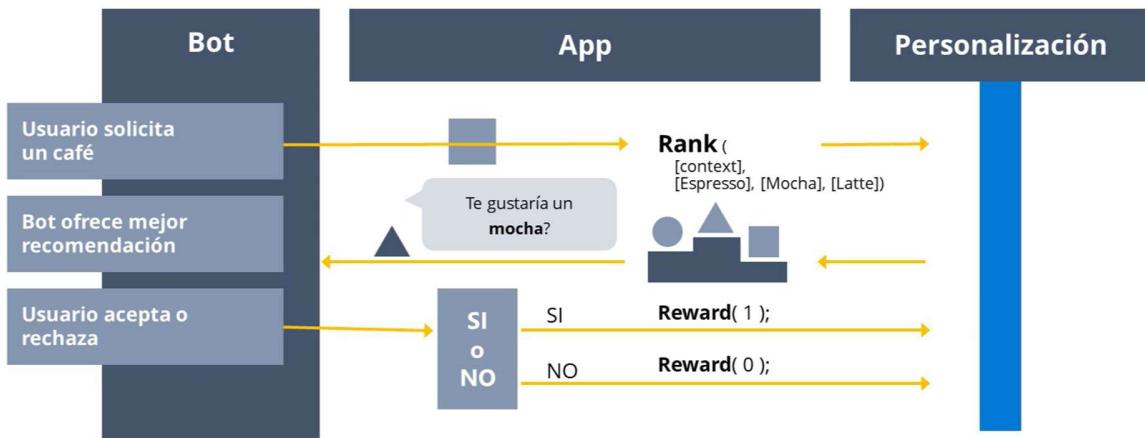
- b. El diseño A y B vienen dado por el analista.
- c. El aprendizaje reforzado se envía una política que indica si es exploración o explotación.



Aprendizaje Reforzado

- 1 Usar modelos que exploten y exploren.
- 2 Registrar la interacción del usuario.
- 3 Encuentra la política y el modelo que se ajusta a la realidad.

- d. El aprendizaje reforzado vendría siendo el A/B Testing corrida muchas veces.
- e. Ejemplo:



Aplicaciones de la visualización en Ciencias de Datos:

1. Aplicación en control de la gestión.
2. Aplicaciones en desarrollo de negocios, ejemplo bandas de artistas preferidas.
3. Existen guías de referencias visuales (buenas prácticas).
4. Existen guías de diseño (buenas prácticas).

Aplicaciones de aprendizajes supervisados: si en el escenario supervisado tenemos datos de premisa, datos de contexto, pero a la vez tenemos etiquetas, conclusiones o algunos datos de las consecuencias del pasado hablaremos de este tipo de aprendizajes.

- El sistema aprende basándose en datos estructurados o no estructurados, previamente clasificados
- **Data:** (datos, etiqueta)
- **Objetivo:** Aprender una función para mapear $x \rightarrow y$
- **Ejemplos:** Clasificación, Regresión, Detección de Objetos, etc.

1. Modelos de fugas de clientes:

- a. Árbol de decisión en industria de seguros.

Conocer los patrones comunes en los casos de
fuga voluntaria de clientes.



2. Identificación de objetos en video: **Mascarilla**



Aplicación de aprendizaje no supervisado:

- 1. Recuperación de cobre.

Aplicaciones del aprendizaje reforzado: Búsqueda rescate

- 1. Hiper Personalización.
 - a. Sugerencia de acciones.
 - b. Lista de reproducción/estaciones.
 - c. Contenido creativo/diseño.
 - d. Campañas de correo electrónico.
 - e. Alertas.
 - f. Sugerir desde un menú.
- 2. Sistemas Autónomos. (Búsqueda para rescate).

Caso en compañía de seguros:

1. Análisis de riesgo.
2. Automatizar el proceso de evaluación de daños.
3. Automatizar la detección de fraude en los reclamos.
4. Cross Selling / Up Selling.
5. Servicio de preguntas y respuestas en tiempo real.
6. Predicción de fugas de clientes:
 - a. Modelo de fuga de clientes:
 - i. Industria seguros.
 - ii. Árbol de decisiones.
 - b. Objetivo de modelo de fuga:
 - i. Conocer los patrones comunes en los casos de fuga voluntaria de clientes.
 1. Fuga dura (cierre formal del contrato) vs fuga blanda (No hay contrato, solo es dejar de usar el servicio, por ejemplo, no ir más al supermercado de cierta marca).
 - c. Metodología:
 - i. Se puede establecer una frecuencia que indique la fuga, por ejemplo, cada cierto mes se promedia y si hay una baja indica fuga.
 - ii. Definir una historia, por ejemplo, 24 meses de historia, uso de promedios históricos móviles 3, 6, y 12 meses, uso de deciles para discretizar valores.
 - d. Árbol de decisiones:
 - i. En este caso, es necesario asignar una probabilidad de fuga a cada cliente y entender cuál es el proceso o patrón detrás de la fuga, es decir, cuáles son los elementos por cambiar del proceso que están produciendo las chances de la fuga de los clientes.

Caso en banca:

1. Modelo de próxima mejor oferta (PMO/NBP): Utiliza algoritmo llamado reglas de asociación. Busca las reglas con mayor fuerza que permita identificar cuál es esta mejor siguiente mejor oferta a un cliente que ya tenemos. Es decir, según el historial del cliente busca la mejor oferta.
2. Las aplicaciones que usan esto son Cross selling, Up selling y recomendador PMO.
3. El objetivo del modelo de mejor oferta es:

Determinar en qué grado, la tenencia de ciertos productos del banco, en combinación con otras variables del cliente, inciden en la propensión a tomar otros productos, de forma de determinar, **cliente por cliente** la próxima mejor oferta.

4. Beneficios de este modelo:

- Enfocar los esfuerzos de ventas con dos miradas (cliente – producto).
- Aumento efectividad de ventas (mejora targeting).
- Calidad de servicio (control hostigamiento).
- Mejora productividad (ejecutivos red de sucursales, Aprovechamiento del inbound).
- Gestión ciclo de vida del cliente.

5. En la construcción de un modelo considerar:

- VARIABLES para considerar: transaccionales internas (tenencias, saldos y uso), Sistema financiero (deudas, líneas aprobadas, instituciones) y Sociodemográficas (antigüedad, edad, género, zona de residencia, renta, etc.)
- Modelo canasta de compra (reglas de asociación): es la creación de reglas, si ocurre algo es probable esto. Estas reglas maximizar dos primicias, el soporte y la confianza.
- Creación de reglas basados en probabilidades: como tomar las reglas e incorporarlas en el proceso de negocio.

6. Ejemplo reglas:

Que combinación es más propensa a: Contratación PAT	Probabilidad
Saldo Prom CC 12m < 100M, Deuda = Existente, FFMM= Existente	83%
Saldo Prom CC 12m < 100M, Línea de Sobregiro= Existente	77%
Saldo Prom CC 12m = 100M- 200M, Deu Sbif < 1MM , Antig=1-2	64%
Saldo Prom CC 12m = 100M- 200M, PAC= Existente	64%
Saldo Prom CC 12m = 100M- 200M, Banca = Preferente	63%

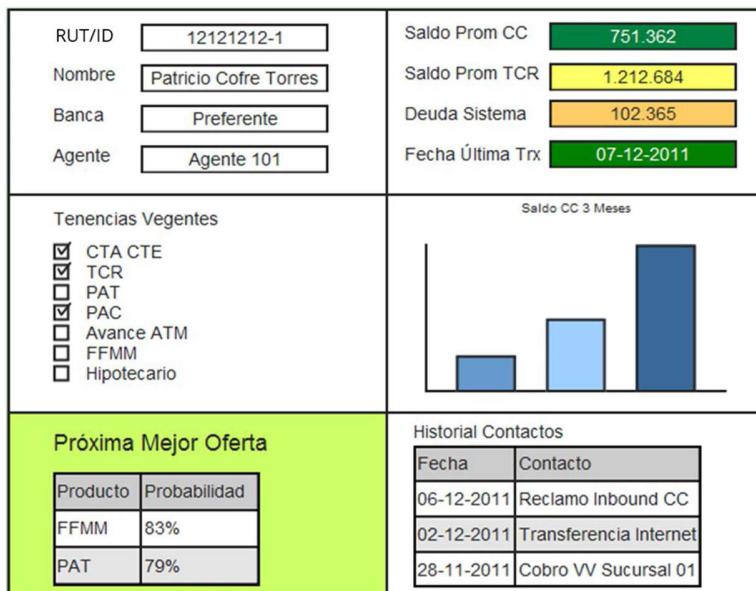
7. Ejemplo reglas aplicado a clientes:

Foco cliente	RUT/ID	Orden	NBP	Probabilidad
Las reglas aplicadas sobre la cartera de cliente permite conocer cliente a cliente las mejores reglas que apliquen en cada caso, para decidir con criterios comerciales cuál ofrecer, o bien ofrecer ambos.	12345678	1	Consumo ATM	83%
	12345678	2	Banca Electrónica	77%
	24680246	1	Hipotecario	74%
	24680246	2	PAC / PAT	69%
	13579135	1	FFMM	79%
	13579135	2	Moneda Extranjera	73%

8. Ejemplo:

Foco producto	RUT/ID	Orden	NBP	Probabilidad
Listar quiénes son los clientes con mayor apetencia de un producto que se desee impulsar.	12345678	1	FFMM	93%
	29292929	2	FFMM	88%
	31313131	1	FFMM	84%
	46545645	1	FFMM	84%
	57957957	1	FFMM	84%
	98765542	3	FFMM	83%

9. Ejemplo:



Caso en retail

1. Se usan los Algoritmos o Sistemas de recomendación (RecSys).

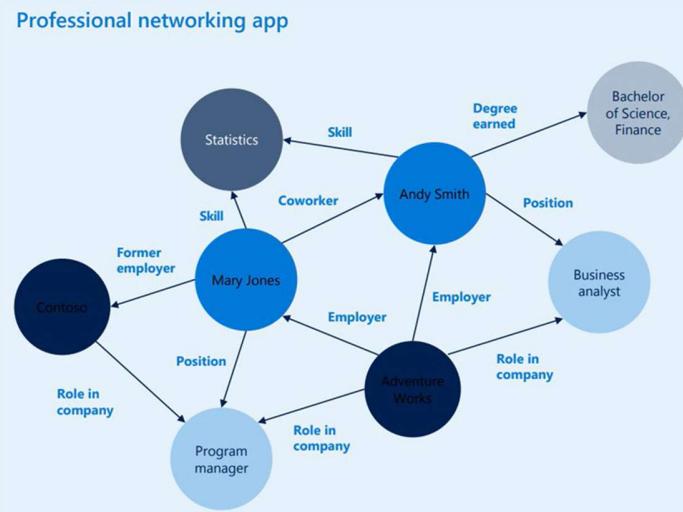
Los sistemas de recomendación corresponden a paradigmas de modelación, capaces de desplegar ítems ordenados en función de una predicción del ranking que un usuario le daría a un ítem **que aún no ha considerado**, o bien no lo ha hecho en la compra actual

2. En RecSys se usan dos familias de enfoque:

- Filtrado basado en contenido: recomendara basado en el análisis de los atributos de los elementos pedidos en el pasado.
- Filtrado colaborativo: las recomendaciones no vienen de los atributos de elementos elegidos, sino que será de usuarios similares a nosotros.

3. Base de datos orientadas a grafos:

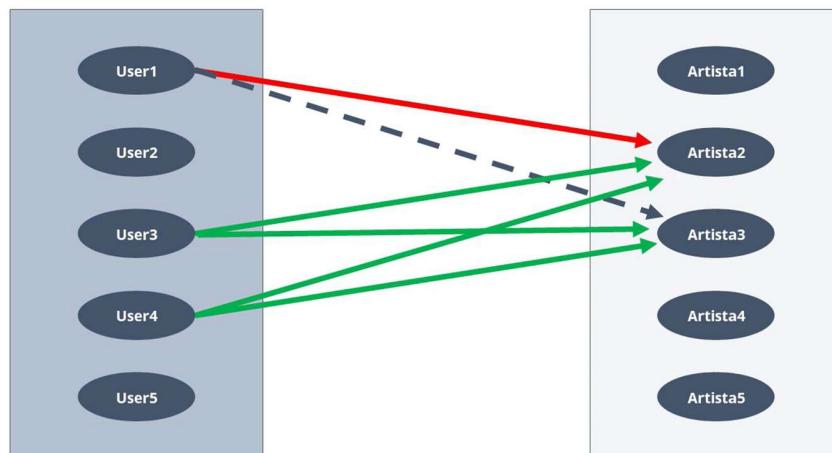
Representación de entidades como **nodos** de un [grafo](#) y sus relaciones como las **aristas** del mismo, de manera que se pueda usar [teoría de grafos](#) para recorrerlo.



4. Las bases de datos orientadas a grafos han resultado útiles:

En sets de datos **evolutivos, interconectados y jerárquicos** han demostrado ser más eficiente que los modelos relacionales. Ejemplos contemporáneos de esto son las **redes sociales** y los sistemas de **recomendación**.

5. Ejemplo recomendación ecommerce para recomendar artistas:



Caso en calidad del aire

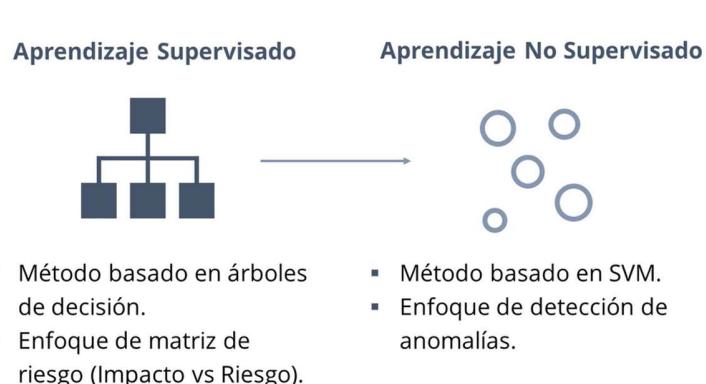
1. Análisis descriptivo: mirada al pasado, al estatus. Permite un primer entendimiento.
2. Análisis predictivo: Se usa el algoritmo de series de tiempo.

Caso en transporte público

1. Desafíos:
 - a. Problema de evasión que en Chile alcanza el promedio de 35%.
2. Actual esquema: Presenta una alta linealidad entre el área cubierta y el personal necesario.
3. Como solución:
 - a. Añadir IA, permitiendo clasificar comportamiento, como la evasión de tarifas y el mal uso de los beneficios.

Caso en compras públicas

1. Objetivo de la iniciativa es optimizar la selección diaria de casos a monitorear por parte de los analistas del observatorio.
 - a. Priorizando los casos en función de la probabilidad de presentar anomalías.
 - b. Entregando antecedentes adicionales que apoyen su toma de decisión.
2. Tipos de modelos:



3. SVM es Support Vector Machine.
4. En detección de anomalías, como sugerencia, para generar modelos usar matrices de riesgos. Ej.



5. Modelo utilizado fue el SVM-One Class Classification:

Objetivo
Definir un clasificador para distinguir las observaciones normales de las anormales.
Requisitos
En el contexto de detección, se necesita un conjunto de datos sin casos anormales como entrada para el entrenamiento del modelo.
Resultados
Para una nueva observación, el clasificador toma valores positivos para "datos normales" y negativos para "datos anormales".

6. Algunas de las variables utilizadas:

Tasa de respuesta a preguntas de los proveedores
Análisis de plazos
Ejemplo Días entre la publicación de respuestas y el cierre del proceso
Ratio de adjudicaciones al mismo proveedor
Fragmentación
Presencia de palabras clave en las preguntas del foro

7. Segundo último variable, las palabras claves fueron, por ejemplo,

- Afecta
- Anomalía
- Auditoria
- Colusión
- Contraloría
- Corrupción
- Flagrante
- Fraude
- Igualdad
- Etc.

8. El proceso posee 4 etapas grandes,

Un **proceso ETL** extrae los **datos** desde los **sistemas transaccionales** de Mercado Público, **procesa** los **modelos de riesgo** y, finalmente, **carga** un **sistema de revisión**.



- Extracción: Tomar a través de los sistemas transaccionales las variables.
- Calculo del modelo: Se corre el modelo y se valida contra la matriz de riesgo.
- Resultados: Se presenta un score y entregar variables, resumen, impacto al analista para que investigue un caso.
- Sistema de revisión: Se presenta todos los casos para que el analista revise.

9. Un sistema de revisión podría ser:

La captura de pantalla muestra la interface de usuario de un sistema de gestión de procesos. La barra superior tiene opciones como "Asignaciones", "Procesos", "Metas" y "Cerrar Sesión".

Los filtros principales incluyen:

- Desde: 30/08/2017
- Hasta: 30/10/2019
- Fase Proceso: Todas
- Tipo de Licitación: Todas
- MEI: Todos
- Estado Gestión: Todos
- Rango COP: Todos
- Destacado: Todos

En la parte inferior, hay tres pestañas: "Licitación" (selecciónada), "Órdenes de Compra" y "Grandes Compras".

La lista de casos es la siguiente:

Destacado	Fecha	Id Proceso	Proceso	Organismo	Encargado	MonitCOP	Score	Reclamos
☆	29-01-2019	66480	Aseo y Cafetería II Región 7	DEPARTAMENTO ADMINISTRATIVO NACIONAL DE ESTADÍSTICA (DANE)		0.0000	0.105	0
☆	29-01-2019	66453	Elementos para Emergencias	FONDO DE PREVISIÓN SOCIAL DEL CONGRESO DE LA REPÚBLICA		0.0000	0.368	0
★	29-01-2019	66387	Nube Privada II	COMISIÓN NACIONAL DEL SERVICIO CIVIL		0.0000	0.226	0

Pantalla inicial

El analista revisa los casos que le han sido asignados, pudiendo priorizar por Score.



Sobre los hallazgos encontrados

El analista debe realizar gestión.
Una vez resuelto, mover al listado
de **hallazgos corregidos**.

Caso industria financiera

Caso observación astronómica

1. Turbulencias ópticas:

Machine Learning

Para predecir con la mayor precisión posible la turbulencia atmosférica para las próximas dos horas en el Observatorio Paranal.

Series temporales de Random Forest(RF) y LSTM

Para entrenar modelos usando datos de 3 años.

- **Problema desafiante:** debido a la turbulencia de la naturaleza estocástica/caótica.
- **Random Forests y LSTM:** uso de modelos de machine learning para capturar patrones y generar predicción

- **Distribución de probabilidades:** absorbe parte de la estocasticidad de alta frecuencia y está más orientada a la operación para la toma de decisiones.
- **Mejores resultados:** el error de predicción es de ~10% en una ventana de 2 horas.

2. Anomalías en Imágenes:

Desarrollar un método para complementar el control visual de las imágenes de calibración realizadas durante el día por expertos. El **método** deberá **identificar imágenes anómalas**.

- Usando técnicas basadas en Deep Learning
- Producir descriptores de imágenes a partir de imágenes sin procesar.

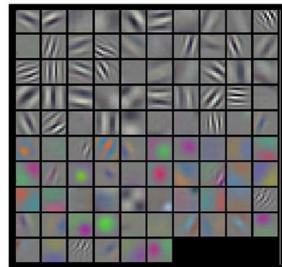
- **Explorar los métodos no supervisados**
para: identificar y clasificar las imágenes en base a sus descriptores:



K Medias
Matriz de similitud
Método Dendrogram

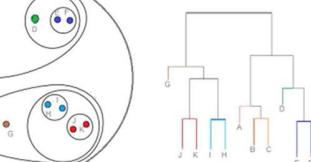
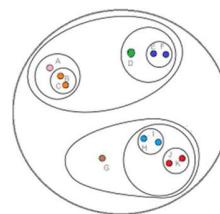
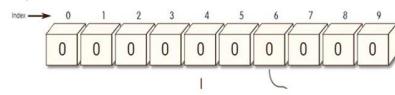
Descriptores que utilizan un ResNet-50 (CNN)

Resolución de las imágenes en bruto
(4096x2048 a 3000x1074)



ración propia

Se calculó el descriptor de imágenes en bruto
ResNet-50 de longitud 2048



Possible detectar anomalías en las imágenes y agrupar los diferentes tipos utilizando técnicas de Deep Learning

Sistema de clasificación y detección de anomalías para complementar el control visual de las imágenes de calibración realizadas durante el día por los expertos.

99% recall en Clasificación de imágenes normales
95% recall en Clasificación de imágenes anómala