

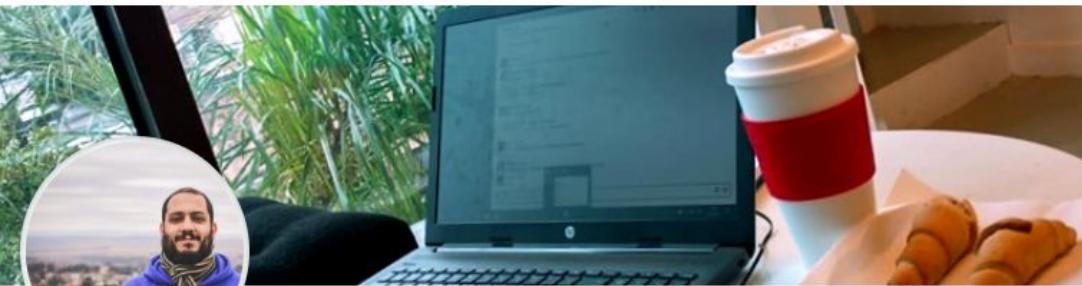
ACADEMICA



Martín Nicolás Gonella · 1er

Lead Engineer & Co-Founder at Adaptic | Data Science Mentor at Acámica
Departamento Capital, Provincia de Córdoba, Argentina

[Más de 500 contactos](#) · [Información de contacto](#)



Juan Manuel Spinelli

Data Engineer en COREBI Analytics & Big Data - Data Science Enthusiast - Provincia de Córdoba, Argentina

Argentina · [Más de 500 contactos](#) · [Información de contacto](#)



Acámica



Universidad Nacional de San Luis



**Acámica es una
academia tecnológica
que acompaña a las personas
en su transformación digital.**



Metodología

Aprendemos haciendo

La creación de productos concretos permite a los/as estudiantes generar soluciones creativas a problemáticas específicas.

Damos feedback que forma

Más allá de certificar que hayan aprendido los contenidos técnicos, buscamos ofrecer un feedback que los ayude a entender sus aciertos y oportunidades de mejora a lo largo del proceso de aprendizaje.

Aprendemos con otros/as

Promovemos clases con multiplicidad de disciplinas en las que se aprende co-creando y colaborando con pares y equipo docente.

Formamos profesionales

Trabajaremos la comunicación, la empatía y la colaboración en distintas instancias para formar egresados que aporten mucho más que saber técnico en los lugares de trabajo a los que ingresen.

**¡Bienvenidos/as al mundo
de Data Science!**



Agenda

- Inmersión en el mundo de los datos: Data Science, Machine Learning, Business Analytics.
- Hay data más allá de los sistemas de recomendación.
- Casos de aplicación.
- Fallando con éxito.
- Actividad práctica: Data Science empieza con preguntas.
- Buenas prácticas de la Ciencia de Datos.
- Reflexiones finales.



INMERSIÓN EN EL MUNDO DE LOS **DATOS**

A nighttime satellite view of Earth from space, showing city lights and auroras.



¿QUÉ ES LA **CIENCIA DE DATOS?**



Es un campo **interdisciplinario** tanto en sus objetivos como en sus metodologías, que busca **convertir datos en conocimiento**, mediante:



DEFINIR

Las preguntas que queremos responder. ¿Qué datos necesitamos para responder esas preguntas?



INVESTIGAR

Se obtienen los datos, se “limpien” y se procede a explorarlos.



ANALIZAR

Los datos obtenidos se analizan con modelos (estadísticos, Machine Learning, etc.). Interpretamos los resultados y transformamos datos en información.



PRESENTAR

Los resultados obtenidos y las conclusiones a las que llegamos. Puesta en producción.



¿Cómo funciona
este proceso en
la práctica?

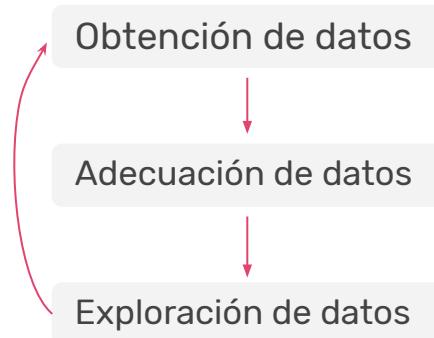
Es un proceso iterativo

DEFINICIÓN

INVESTIGACIÓN

ANÁLISIS

PRESENTACIÓN



Es un proceso iterativo



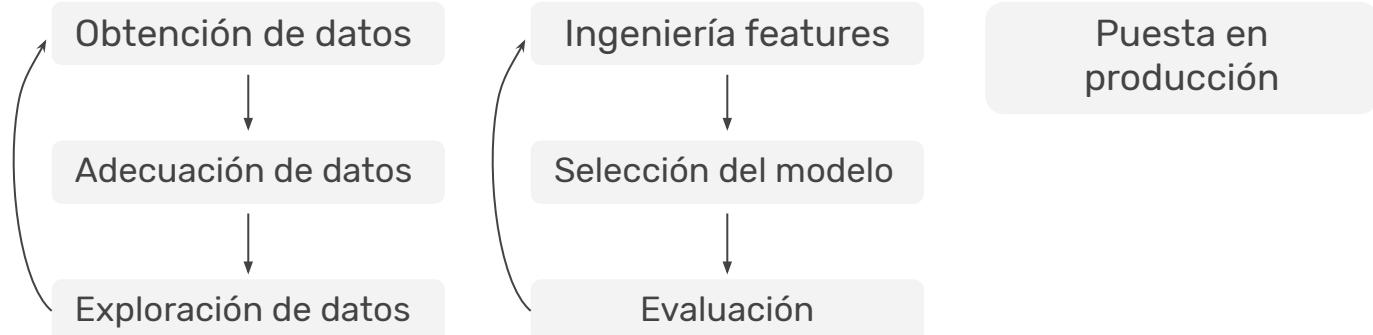
Es un proceso iterativo

DEFINICIÓN

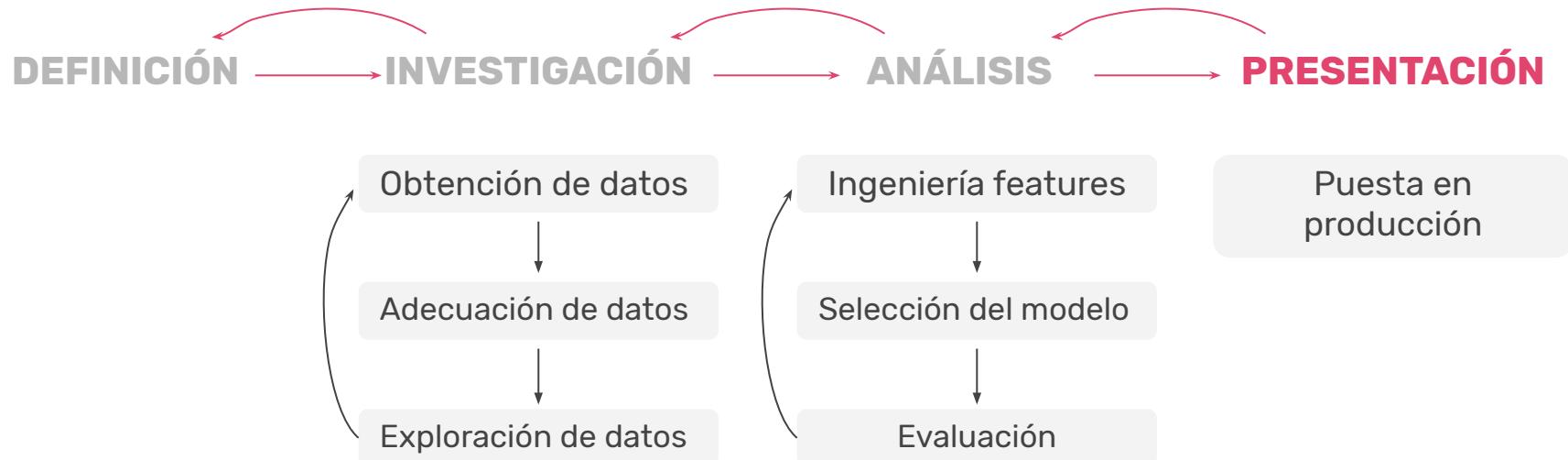
INVESTIGACIÓN

ANÁLISIS

PRESENTACIÓN



Es un proceso iterativo







La clave está en hacerse las
preguntas correctas!



¿Puedo **automatizar** la toma de decisión de mis procesos?

¿Qué puedo hacer para **aumentar mis ventas**?

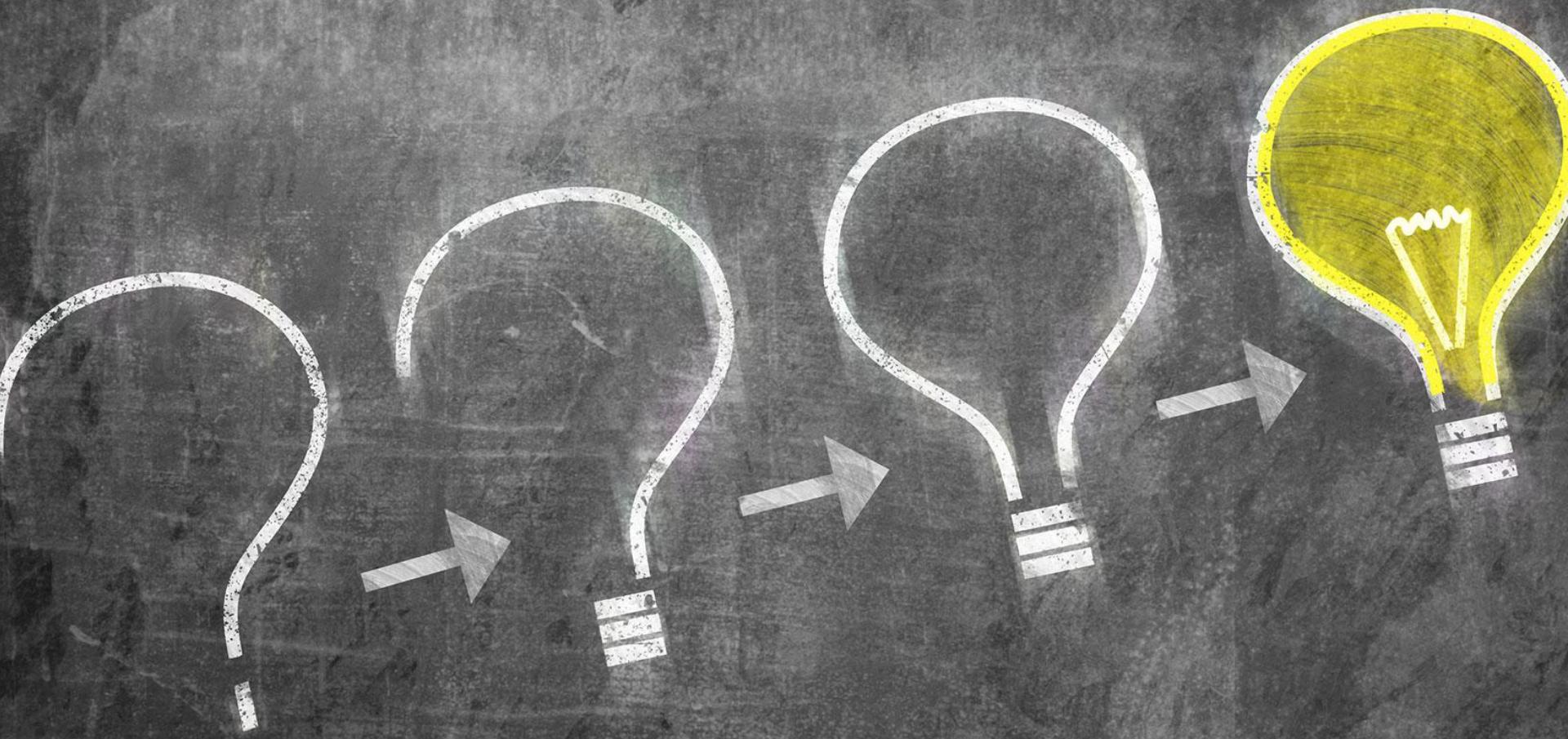
¿Cómo me **anticipo a problemas** en mi cadena de producción?

¿Qué hago con los **años de datos almacenados** que tiene mi empresa?

¿Qué puedo hacer para **mejorar el engagement** con mis clientes?

¿Por qué está disminuyendo la **fidelidad de mis clientes**?

La **seguridad** de mi empresa depende de la atención del personal de vigilancia ¿Se puede mejorar?





***“Los datos son el
nuevo petróleo.”***

- Martin Hilbert



*~~“Los datos son el
nuevo petróleo.”~~*

- Martin Hilbert

***“Toda empresa es una
empresa de datos.”***

Y si no nos creen a nosotros...

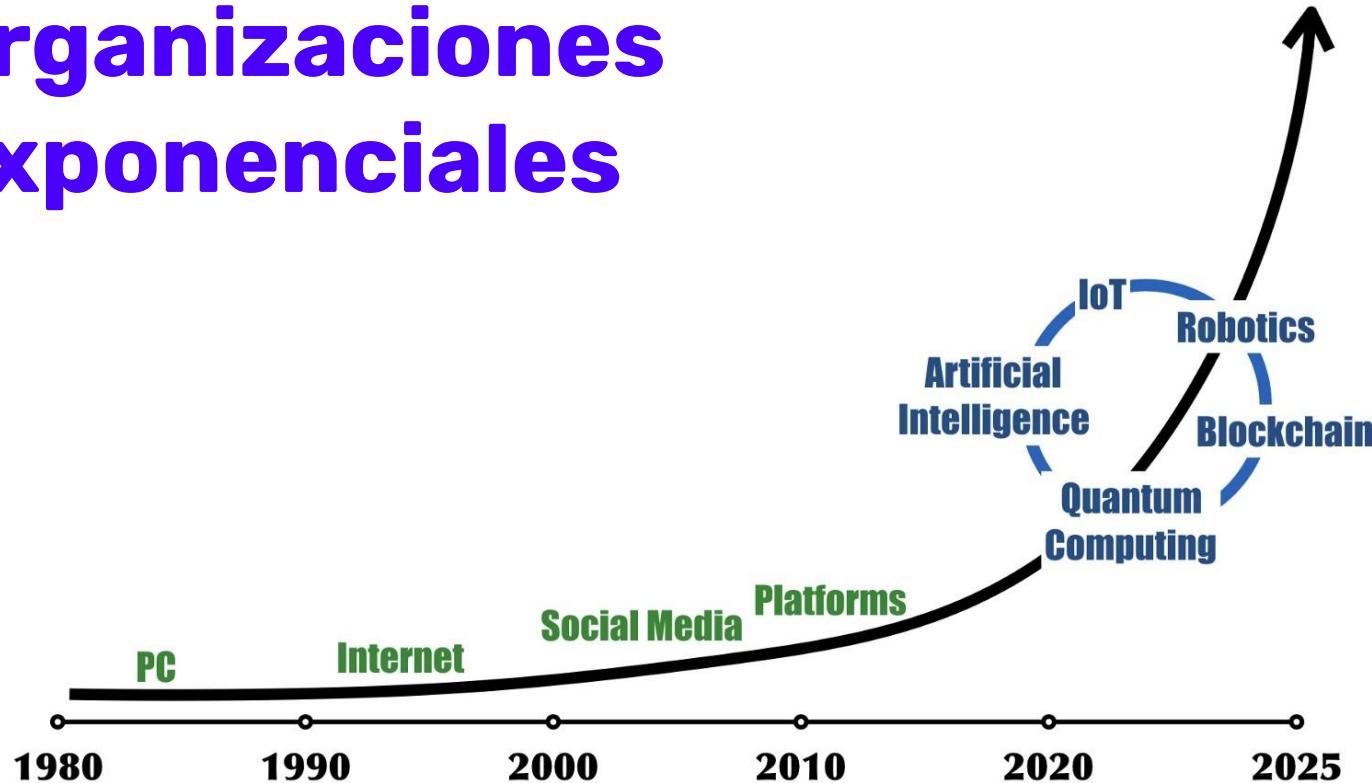


Data is key to Starbucks, which includes a head of Global Strategy, Insights and Analytics as part of its executive leadership team. According to Starbucks, this function uses **“methodologies ranging from ethnography to big data analytics [...] that helps support Starbucks pricing strategy, real estate development planning, product development, trade promotion optimization and marketing strategy.”**

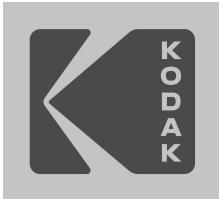
HAY DATA **MÁS ALLÁ** DE LOS SISTEMAS DE RECOMENDACIÓN

Negocios	Mejoras en los productos y servicios. Patrones de comportamiento. Métricas. Previsiones. Atención al cliente, mejoras de engagement, predicciones de abandono, etc.
IoT & Industria 4.0	Detección de anomalías, predicción de fallas. Análisis de grandes volúmenes de datos provenientes de miles de sensores. Mejoras en la toma decisiones, etc.
Agricultura & Ganadería	Prescripciones de siembra, riego y fertilización. Detección y diferenciación de malezas y cultivos. Detección de enfermedades en cultivos y animales. Trazabilidad de rindes con imágenes satelitales, GIS+AI, etc.
Entretenimiento	Sistemas de recomendación. Generación automática de subtítulos. Experiencias de usuarios personalizadas, Marketing y publicidades personalizadas. Búsquedas optimizadas, etc.
Finanzas	Bots Traders en monedas/criptos. Subastas de publicidad. Análisis de riesgo crediticio. Detección de fraudes. Modelados económicos. Asistencia al cliente, etc.
Salud	Mejoras en el diagnóstico de enfermedades. Detección muy temprana de enfermedades. Tratamientos personalizados contra enfermedades como el cáncer. Cirugías asistidas por robots, etc.

Los datos permiten crear organizaciones exponenciales



1996



Market Cap: \$28B
Employees: 140.000

2012



Market Cap: Bankrupt
Employees: 17.000



Market Cap: \$1B
Employees: 13

Source: Peter Diamandis - Singularity University

Cambia, todo cambia. Nuestro mundo hoy...



U B E R

La empresa de taxis
más grande del
mundo no es dueña
de ningún vehículo.



La empresa de social
media más popular
del mundo no crea
ningún contenido.



La empresa de
alojamientos más
grande del mundo
no posee ningún
inmueble.



El retailer más
grande del mundo
no tiene inventario.

DATA SCIENCE

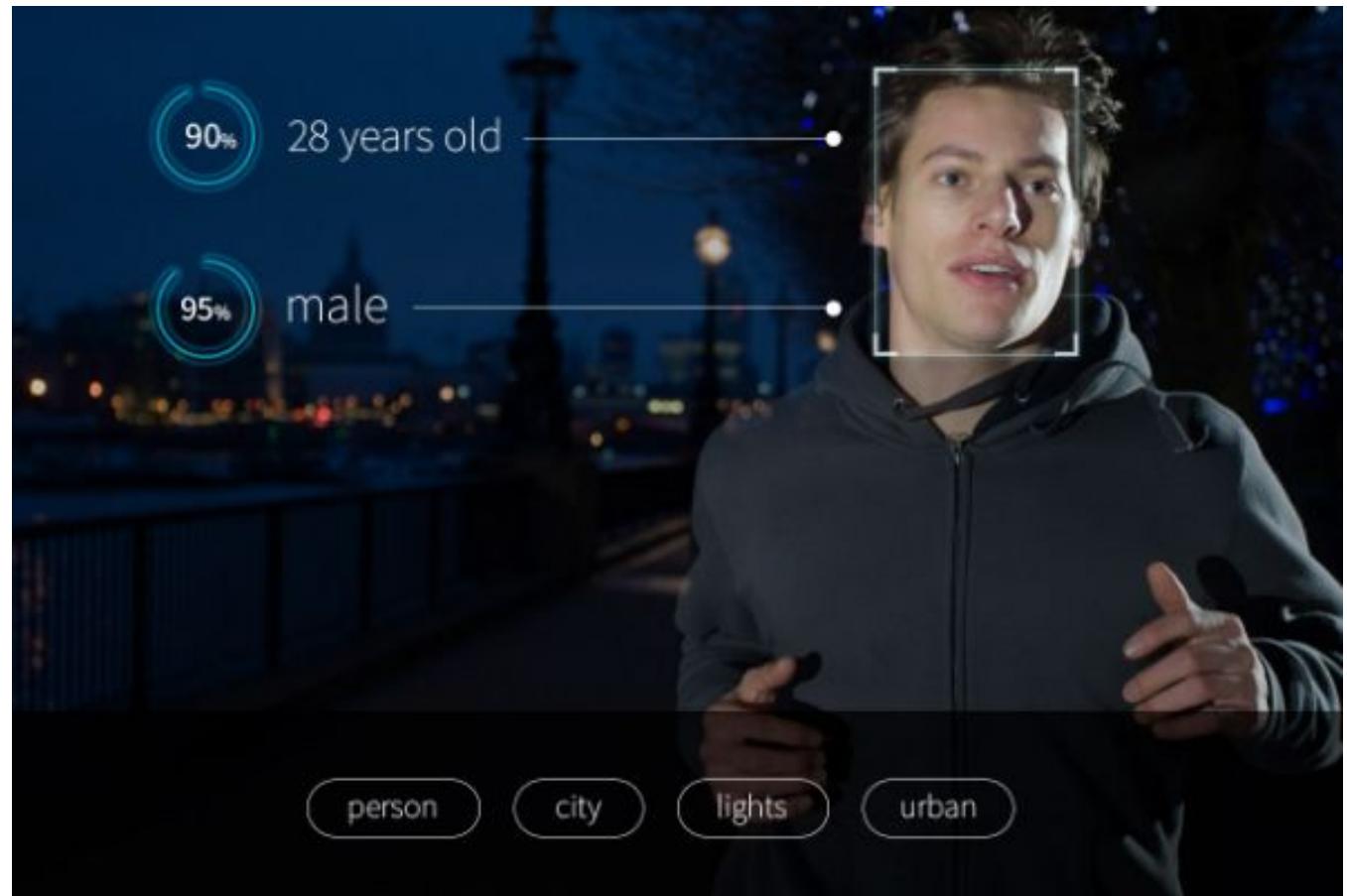
EN LA VIDA REAL

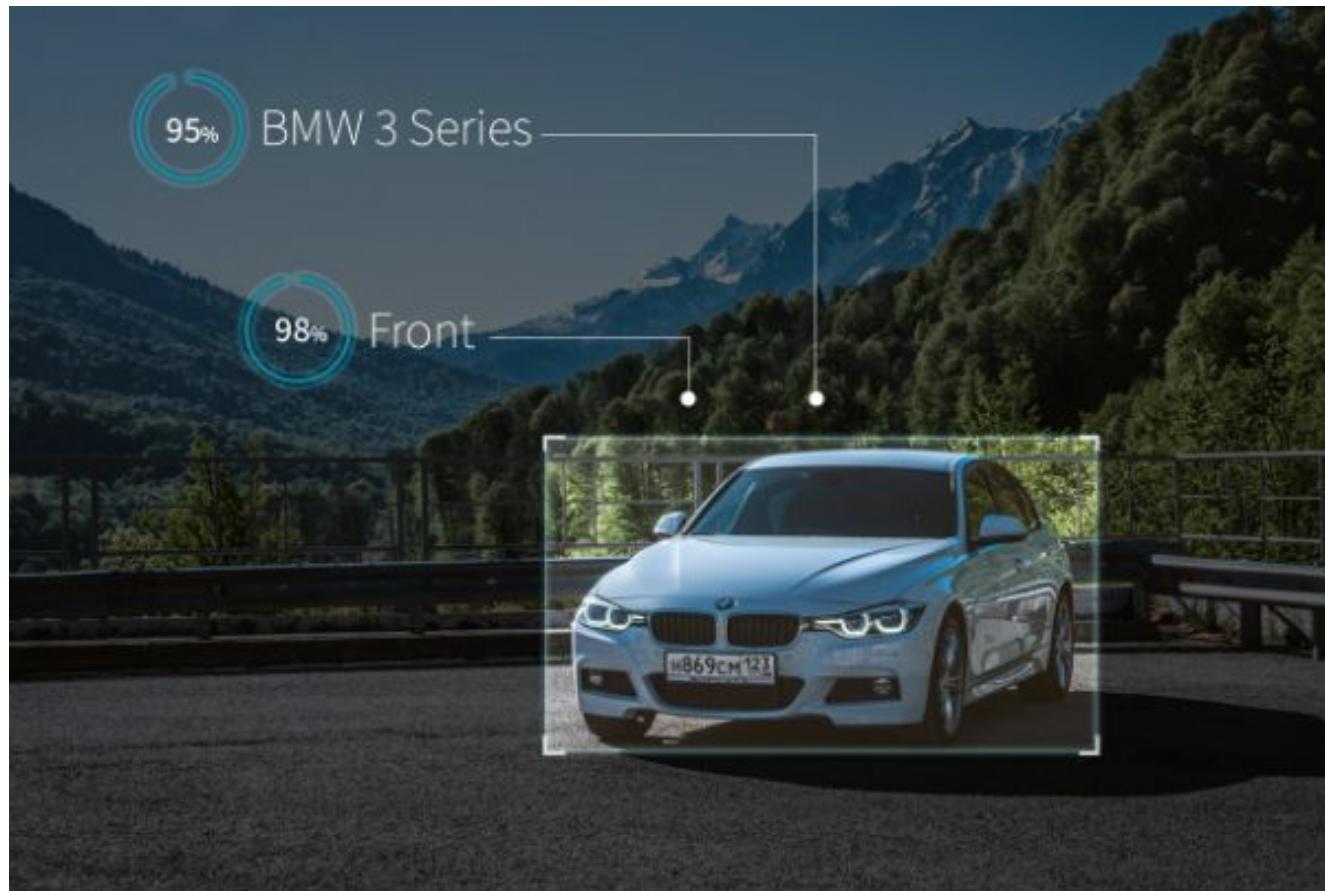
Ph. by Mauro Mora on Unsplash

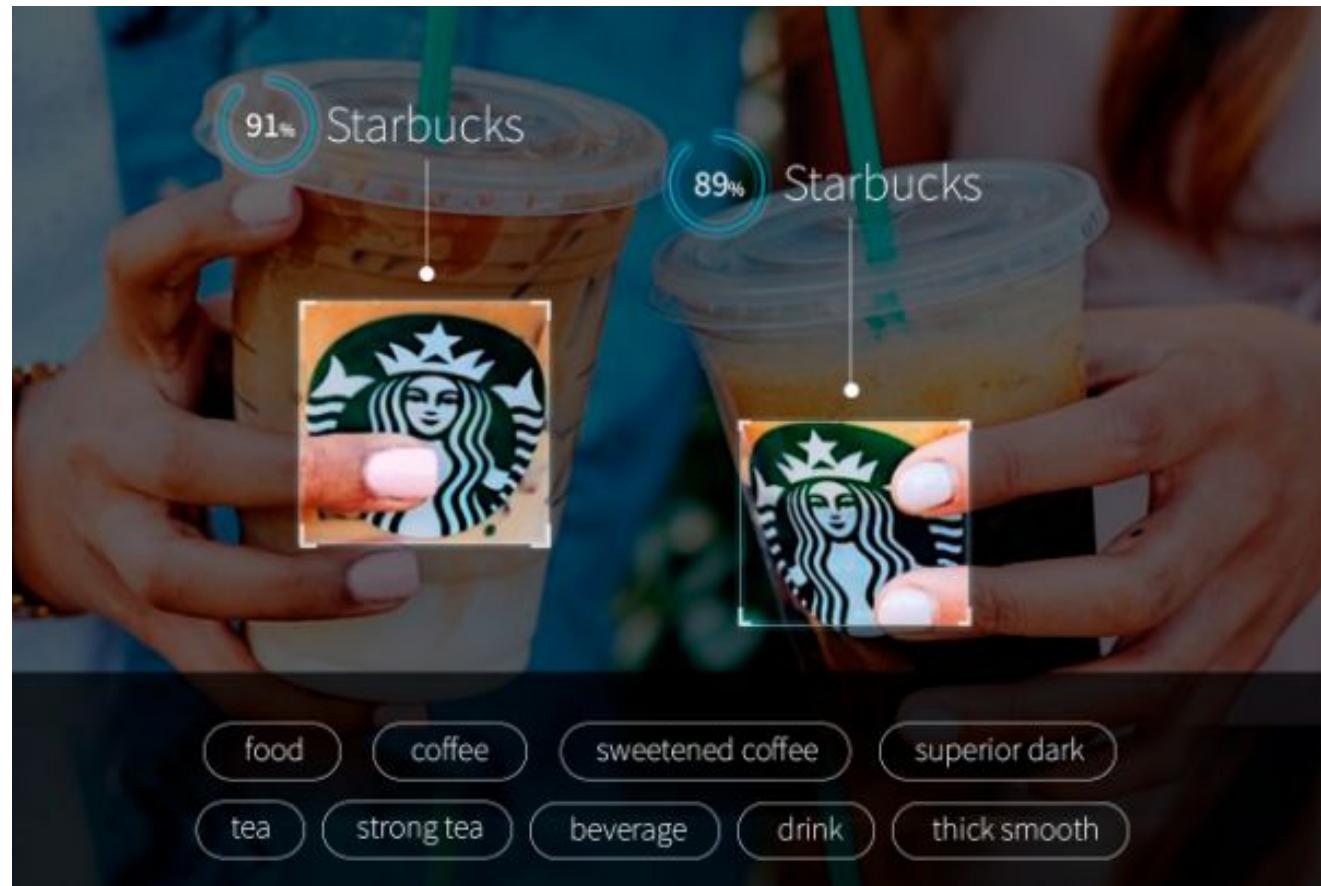


VISUALLY SIMILAR PRODUCTS

Computer Vision









00:00:00

cart



Fintech



brubank

¿Quiénes somos?

Precios

Seguridad

Ayuda

Robo o extravío

No somos la app de un banco.

**Somos un banco
en una app.**

Bajátela y sé de los primeros en tener una cuenta.



Ventas, Mkt y Publicidad

 <small>Imagenes ilustrativas</small> \$ 15.999 30% OFF Envío gratis Pc Armada Gamer Amd A8 9600 X10 Nucleos Video R7 Hdmi Ssd240	 \$ 11.499 23% OFF Envío gratis Pc Armada Slim Intel Amd Dual Core Hd 1tb Minecraft Lol Hdmi	 \$ 28.102 Hasta 18 cuotas sin interés Llega gratis mañana  FULL Pc Completa Gamer Amd Ryzen 5 2400g 1tb 8gb Ram
 \$ 9.999 23% OFF Envío gratis Pc Computadora Amd Intel Dual Core Vga Hdmi 4gb Ddr3 Netflix	 \$ 18.999 24% OFF Envío gratis Pc Armada Gamer Amd A8 9600 X10 Video R7 8gb Ram Hdmi W10 64	 \$ 14.999 21% OFF Envío gratis Pc Gamer Amd A6 7480 4ghz X6 8gb Ram Ssd240 Hdmi Radeon R5



Shipping Delivery Time prediction

BUSINESS CASE

A solution to provide a useful delivery date prediction which has a direct impact on customer experience.

Delivery Promise tuning
Ability to manage the trade-off accuracy/late delivery.

Handling time buffers
Based on item/provider record.

Shipping delays

Customer experience

Content Personalization

AI Drives Relevant Dialog with Customers

**GARBARINO**

Heladera con Freezer Patrick
HPK151M11N Negro 388 Lts.
Antes: \$52.725
iAprovecha \$44.999!

**GARBARINO**

Heladera con Freezer Patrick
HPK151M11S Plateado
Antes: \$50.936
iAprovecha \$43.999!

**GARBARINO**

Heladera No Frost Electrolux
DFN3500P Plata
iAprovecha \$42.999!

**GARBARINO**

Heladera con Freezer Patrick
HPK136A00N Negro 300 Lts.
Antes: \$42.672
iAprovecha \$34.999!

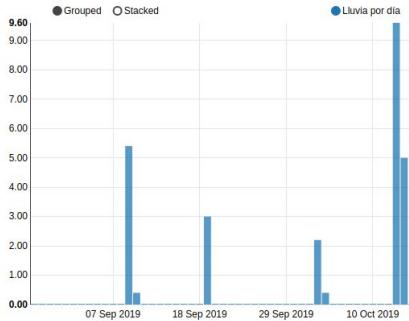
**GARBARINO**

Heladera No Frost GE
Appliances RGA1130YGRU
Antes: \$54.907
iAprovecha \$45.999!

EL CABURE 1 #543

[Volver](#)

Internet of Things



Fecha	Lluvia
14/Oct/2019	5,00
13/Oct/2019	9,60
04/Oct/2019	0,40
03/Oct/2019	2,20
19/Sep/2019	3,00
10/Sep/2019	0,40
09/Sep/2019	5,40



LLUYAS DE TOMO
FECHA INICIO DE SESIÓN: 3 de Julio de 2019 a las 12:00
Lista de Trackers

Datos Crudos → Datos Procesados → Datos Limpios & Procesados →



Nº	Marca de tiempo	Latitud	Longitud
1	2019-10-15 10:52	-31,772700	-63,817900
2	2019-10-15 08:52	-31,772800	-63,817900
3	2019-10-15 06:50	-31,772800	-63,817900
4	2019-10-15 04:50	-31,772800	-63,817900
5	2019-10-15 02:48	-31,772900	-63,817800
6	2019-10-15 00:48	-31,772800	-63,817900
7	2019-10-14 22:46	-31,772800	-63,817900
8	2019-10-14 20:46	-31,772800	-63,817900
9	2019-10-14 18:44	-31,772700	-63,817800
10	2019-10-14 16:44	-31,772800	-63,817900
11	2019-10-14 14:42	-31,772800	-63,818000
12	2019-10-14 12:42	-31,772800	-63,817900
13	2019-10-14 10:40	-31,772800	-63,817900
14	2019-10-14 08:40	-31,772800	-63,817900
15	2019-10-14 06:38	-31,772900	-63,818000
16	2019-10-14 04:38	-31,772800	-63,817900
17	2019-10-14 02:36	-31,772800	-63,818000
18	2019-10-14 00:36	-31,772800	-63,817900
19	2019-10-13 22:34	-31,772800	-63,817900
20	2019-10-13 20:34	-31,772800	-63,817900

Internet of Things

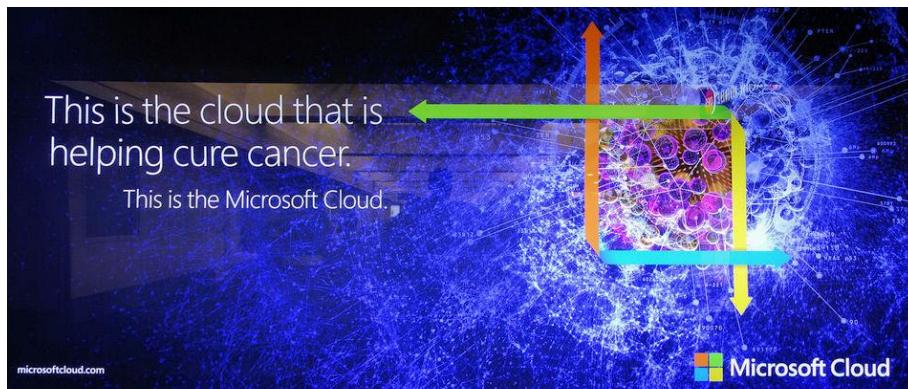


Salud

IBM Watson Oncology Advisor

Treatment Plan	Confidence	Patient Preferences Match
Treatment plan 1 Bisokinol, Olaparib, Cetuximab, Pemetrexed, Bevacizumab	95%	Acceptable match with patient preferences
Treatment plan 2 Systeamic Chemotherapy, Pembrolizumab, Radiation	45%	Unacceptable match with patient preferences
Treatment plan 3 Systeamic Chemotherapy	8%	Preferred match with patient preferences
Radiation and Surgery are unlikely to be appropriate.		

Treatment Options IBM WATSON



KHEIRON MEDICAL TECHNOLOGIES

Home Team Mia RSNA 201

Our mission is to help radiologists detect breast cancer earlier using deep learning.

Watch Video ▶

Your.MD

Sym

Find Your Health

Your.MD is a free service to help you manage your health issues and find the path to a healthier you.

Join the 3,000,000 people using Your.MD

Entretenimiento

The Spotify interface displays a "Made For You" section with two main cards: "Uniquely yours" featuring "On Repeat" and "Repeat Rewind". Below these are "Your Daily Mixes" cards for "SONIDO DE FONDO PARA 16EROS" and "FONDO DE FONDO PARA 17EROS". The left sidebar shows sections like Home, Browse, Radio, and various playlists such as "Made For You", "Recently Played", and "TemakenesClasicos".

The YouTube interface shows a search bar at the top with the query "que realmente podemos dar vuelta la historia". Below it is a "Recomendados" section with three video thumbnails: "ROHAN MUSIC & AMBIENCE" (2:00:01), "How to Become a Data Analyst in 2019" (8:31), and "20 cosas que NO VISTE en JOKER" (14:13). The left sidebar includes links to Inicio, Tendencias, Suscripciones, Biblioteca, Historial, Ver más tarde, Videos que me g... (with a loading icon), and Les luthiers - Vieja... (with a loading icon).

The Netflix interface features a "Novedades de esta semana" section with thumbnails for "CÓMO VIVIR CONTIGO MISMO", "LAVANDERÍA", "diecisiete", "ELI", and "HERIDAS". Below this is a "Nuestra selección para Martin" section with thumbnails for "CÓMO VIVIR CONTIGO MISMO", "BoJack Horseman", "BIG MOUTH", "FRIENDS", and "DARK". The top navigation bar includes links to Inicio, Programas, Películas, Agregados recientemente, Mi lista, and user profile icons.

Agricultura & Ganadería



TECHNOLOGY SERVICE TEAM CONTACT US LOGIN

WE ARE THE NEW AGRICULTURE



SOIL SAMPLES

We collect soil samples from the farm and adjust our algorithm to the specific conditions of each plot of land.



BIG DATA

We automatically collect and analyze all the necessary climate and satellite data to keep our water balance algorithm running 24/7.



DECISION SUPPORT

Our web and mobile application suggests when and how much water to irrigate with, based on the irrigation strategy chosen by the farmer.



FALLANDO CON ÉXITO

EN DATA SCIENCE

MODELO CHURN DE CLIENTES

Empresa de Seguros

Descripción

El **área de IT** se propone crear el primer proyecto de ciencia de datos en la compañía: un **modelo de churn** para implementarlo en el **área de patrimoniales** para evitar la pérdida de clientes

Desafíos superados

1. Implementación de nuevas tecnologías.
2. Pocos datos de personas que abandonan la compañía respecto a quienes se quedan en la compañía.
3. Consolidación de información, múltiples fuentes.

MODELO CHURN DE CLIENTES

Empresa de Seguros

Fallas

- No se involucró a un área clave
- La selección de datos se efectuó en base al conocimiento del área de IT sobre la problemática del negocio



Soluciones

1. Se cambió el Product Owner por un referente de negocio.
2. Se realizó un trabajo intensivo de ingeniería de variables.



VIDEO ANALYTICS

Empresa Financiera

Descripción

En plena transformación digital y pensando en migrar los canales de atención presencial a canales digitales, desde el área de customer experience quería medir cuántas personas se presentaban en las sucursales, con una apertura por Edad, Sexo, Franja Horaria y Recurrencia.

Desafíos superados

1. Implementación 300 sucursales -> Hardware Económico.
2. Tráfico de Red -> Procesamiento Híbrido (Local/Cloud).
3. Medir recurrencia -> Comparar rostros.
4. Regulaciones -> No se guardaban fotos de personas.

VIDEO ANALYTICS

Empresa Financiera

Fallas

No se identificó claramente la **COMPETENCIA** de la solución, en paralelo a la solución se estaba desarrollando un turnero, para el cual solicitaba el RUT, con esos datos se recolectaban las métricas que se buscaban mediante la solución de video.



Soluciones

Se cambió el caso de uso para detectar expresiones “sonrisas” de clientes en locales. Usando esa variable como índice de satisfacción en las gestiones.



PREDICCIÓN FUTURA DE STOCK

Empresa Productora de Bebidas

Descripción

El **área de comercial**

propone construir un modelo que realice una predicción de fuera de stock para pequeños almacenes.

Permite optimizar las ventas y la logística en la distribución.

Desafíos superados

1. Incorporación de datos externos.
2. Georreferenciar locales, validar direcciones.
3. Disponibilizar datos evitando dependencias de IT.

VIDEO ANALYTICS

Empresa Productora de Bebidas

Fallas

Mal concebido el flujo de operaciones: Disponibilidad de datos etiquetados de cuando un almacén se quedaba sin stock, no había una correcta carga por parte de los vendedores.



Soluciones

Se cambió el foco de la predicción, se orientó a la predicción de demanda de comercios minoristas.



PREDICCIÓN DE FRAUDE

Empresa Pública de Seguros Médicos

Descripción

La dirección general de la empresa se propone implementar un modelo de Machine Learning para predecir fraudes en prestaciones médicas, tanto en la atención por parte de centros médicos, como en la venta de medicamentos.

Desafíos superados

1. Gran volumen de datos.
2. Mala calidad de datos.
3. Sistemas origen obsoletos (1600 tablas sin integridad referencial).

PREDICCIÓN DE FRAUDE

Empresa Pública de Seguros Médicos

Fallas

No hubo acuerdo en determinar qué es fraude y que no, lo que imposibilitó el correcto etiquetado de las clases. Si hay excepciones, no es fraude, por más que haya muchos datos, si no es posible el etiquetado, no es posible aplicar técnicas de clasificación.



Soluciones

Se cambió el caso de uso para detectar anomalías. Aún se sigue trabajando en la definición de fraude.



PROPENSIÓN A SUSCRIPCIÓN PREMIUM

Site digital de noticias

Descripción

El **área de IT** decidió implementar un modelo de ML para determinar la propensión de los lectores digitales a **suscribirse a la membresía premium**. La implementación se integra con el Muro de Pago para ofrecer la suscripción online.

Desafíos superados

1. Gran volumen de datos.
2. Variedad de formatos información.
3. Interacciones sistema-sistema.
4. Manejo de datos externos.

PROPENSIÓN A SUSCRIPCIÓN PREMIUM

Site digital de noticias

Fallas

- El modelo funcionó correctamente los primeros 3 meses. Luego gradualmente **fue perdiendo eficiencia.**
- No se designó a un área de negocio que realice un soporte evolutivo del modelo, el contexto cambió, y el modelo quedó obsoleto..



Soluciones

- Se incorporaron perfiles de ciencia de datos en el negocio para que pueda tomar el soporte evolutivo del modelo.
- Actualmente es uno de los pilares para la suscripción.



Actividad:

MACHINE LEARNING CANVAS

Generando una propuesta de valor

- Cuando comenzamos un proyecto de ML existe la tentación de tomar datos, aplicar algoritmos, calcular métricas y obtener valor en todo tipo de formatos incluso cuando estas métricas no tienen interpretación en el dominio del negocio.
- Estamos tan centrados en estas cosas que perdemos la noción del panorama en general.

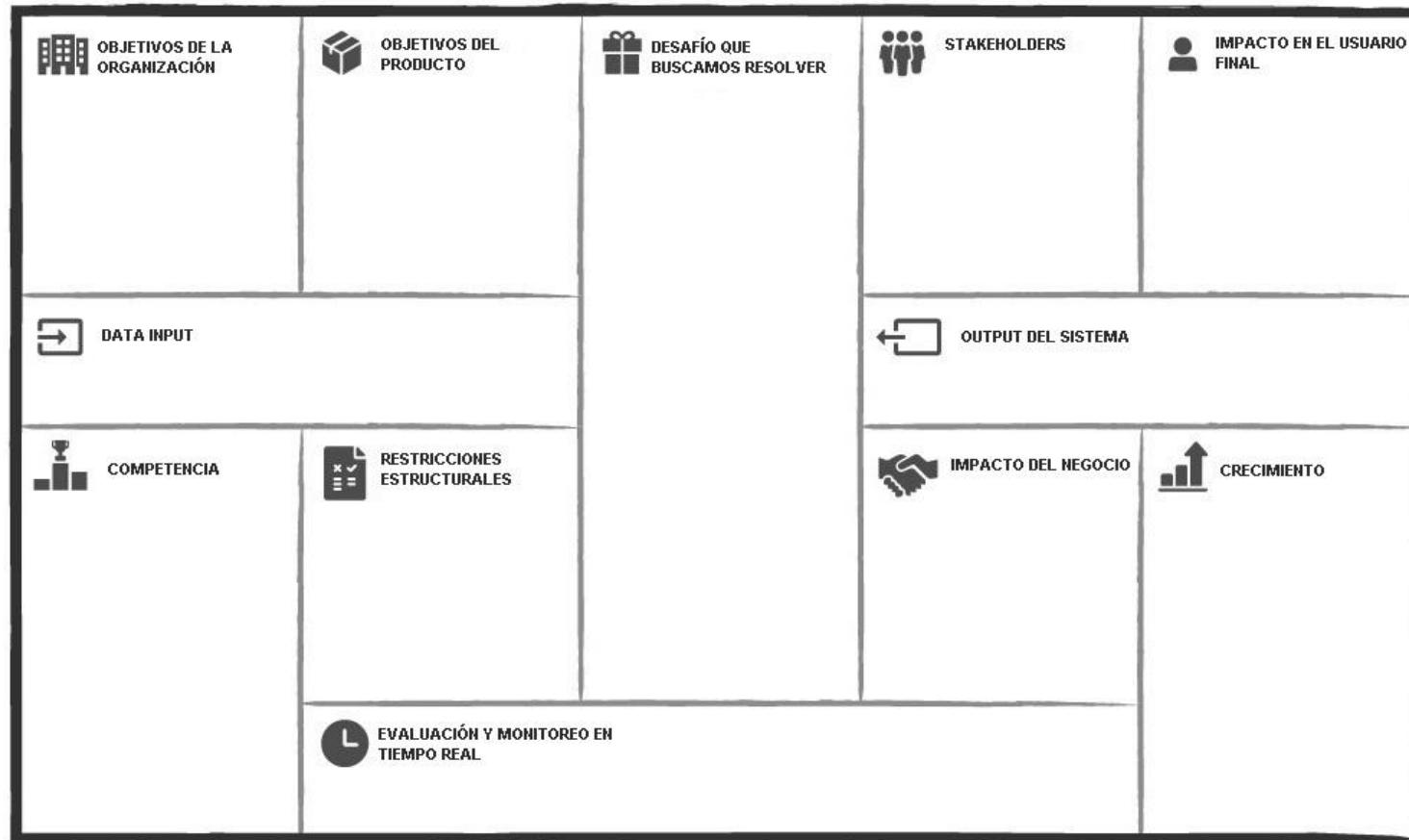




Usar un “Canvas” o “Lienzo”

- En estos casos, una buena práctica es usar un canvas o lienzo.
- Un "lienzo" es solo un gráfico visual para describir un objeto complejo.
- Cada aspecto clave tiene su propio "bloque", y los bloques se organizan en el gráfico de una manera que tiene sentido visual.
- El lienzo es lo que nos acercará a la implementación, con la confianza de que estamos resolviendo el problema correcto y que estamos eligiendo las herramientas adecuadas.

Machine Learning Canvas



Crearemos un Machine Learning Canvas

El objetivo de esta actividad es darles una herramienta que les facilite el proceso de creación y la futura explicación de una idea a personas que pueden ser o no del rubro en el que ustedes se desenvuelven.



#1. Desafío que buscamos resolver



DESAFÍO QUE
BUSCAMOS RESOLVER

¿Qué y cómo?

La tarea que queremos realizar. Qué problema vamos a resolver implementando el modelo.

En esta parte no es solo identificar el problema, si no cuál es el desencadenante.

Algunas preguntas que nos podemos hacer:

- **¿Cuál es el problema que queremos solucionar?**
- **¿Por qué empieza este problema?**
- **¿Qué vamos a hacer y cómo?**

#2. Objetivos de la organización



La búsqueda de la organización o compañía detrás de esta inversión.

Los objetivos organizacionales plantean un horizonte de trabajo, nos permiten construir trazabilidad desde la propuesta hacia lo que moviliza a la compañía, sea rentabilidad, crecimiento en los ingresos o posicionamiento.

Algunas preguntas que nos podemos hacer:

- **¿Qué se busca con esta inversión?**
- **¿Qué ganamos?**
- **¿Seguir adelante con esta idea genera algún conflicto con algún objetivo de la organización?**

#3. Objetivos del producto



Qué objetivos busca la propuesta de valor que estamos planteando. En este espacio se busca que los objetivos no estén expresados de manera técnica sino expresarlos de una forma en la que todos los participantes entiendan sin tener que recurrir a un glosario, documentación, etc.

Algunas preguntas que nos podemos hacer:

- **¿Qué métricas vamos a perseguir?**
- **¿Buscamos que el producto sea eficaz, eficiente o efectivo?**

#4. Stakeholders



Los actores directamente involucrados en el uso y consumo del producto. En esta instancia debemos entender el impacto de la solución en las tareas de nuestro actores.

Algunas preguntas que nos podemos hacer:

- **¿Quiénes y cómo van a usar el producto?**
- **¿Qué ganan o pierden?**
- **¿Qué conflictos pueden surgir entre los actores al momento de aplicar esta solución?**

#5. Impacto en el usuario final



Saber cómo impactamos al usuario es un punto de salida de toda la propuesta, y por eso su ubicación en el canvas. Cómo impacta en el usuario un malfuncionamiento es otro de los puntos claves.

Algunas preguntas que nos podemos hacer:

- **¿Qué tiene o qué puede hacer el usuario final ahora?**
- **¿Cómo cambian sus tareas?**
- **¿Qué valor agregado generamos?**

#6. Data input



Las probabilidades de éxito de un proyecto de Machine Learning están asociadas a la calidad y disponibilidad de datos relevantes. Este factor debe ser explícito al contar la historia, o caemos en el pensamiento mágico de un algoritmo que, con pocos datos de calidad dudosa puede proveer un resultado extraordinario.

Algunas preguntas que nos podemos hacer:

- **¿Qué datos tenemos?**
- **¿Qué datos quisiéramos o esperamos tener?**

#7. Output del sistema



- Información
- Un número concreto
- La toma de una decisión

Esta forma de definir el formato de salida va a facilitar el análisis y disminuye el retroceso.

Algunas preguntas que nos podemos hacer:

- **¿Qué genera la solución y en qué formato?**

#8. Competencia



El análisis del ambiente competitivo no consiste sólo en identificar otras opciones a nuestra solución, sino también reconocer, qué es lo que esos competidores ya han instalado. Algunos competidores pueden incluso no pertenecer al dominio.

Algunas preguntas que nos podemos hacer:

- **¿Cómo resuelven este problema los competidores o la industria en general?**
- **¿Actualmente tenemos algún sistema o proceso que compita contra esta solución? Si es así, ¿En qué nos beneficiará aplicar esta solución?**

#9. Restricciones estructurales



RESTRICCIONES
ESTRUCTURALES

En este bloque vamos a dejar evidencia de los costos que se asumen construyendo este producto: los actuales y los futuros.

Algunas preguntas que nos podemos hacer:

- **¿Qué riesgos o costos futuros queremos asumir o no de manera consciente?**
- **¿Contamos con la infraestructura necesaria?**
- **¿Vamos a trabajar con nuestros servidores o podemos migrar a la nube?**
- **¿Tenemos gente capacitada para que pueda utilizar el modelo y le de mantenimiento?**
- **¿Vamos a capacitar a alguien o contrataremos soporte?**

#10. Impacto en el negocio



¿Cuál será el impacto de nuestra solución en el negocio?
¿Es un ahorro o una ganancia?
¿Es monetizable?
¿Es una vitamina, algo para ganar una oportunidad o un analgésico, algo para resolver una necesidad?

Algunas preguntas que nos podemos hacer:

- **¿Qué va a cambiar en el negocio?**
- **¿El cambio es concreto y directo o indirecto y difuso?**

#11. Crecimiento



Alejándose de los stakeholders considerados al inicio y del impacto en las tareas de los usuarios finales específicos, no es raro que un producto intensivo en Machine Learning tenga potencial de expansión y crecimiento hacia otros stakeholders y otras tareas de los usuarios finales.

Algunas preguntas que nos podemos hacer:

- **¿La solución puede generar a futuro más valor para otras áreas?**
- **¿Vamos a poder escalar esta solución para abordar otros temas en el futuro?**

#12. Evaluación y monitoreo en real-time



Compartir entre el equipo, técnico o no, cuáles de todos los componentes mencionados arriba van a ser los factores a observar para que la Propuesta de Valor sigue siendo un eje coherente dentro del Canvas.

Algunas preguntas que nos podemos hacer:

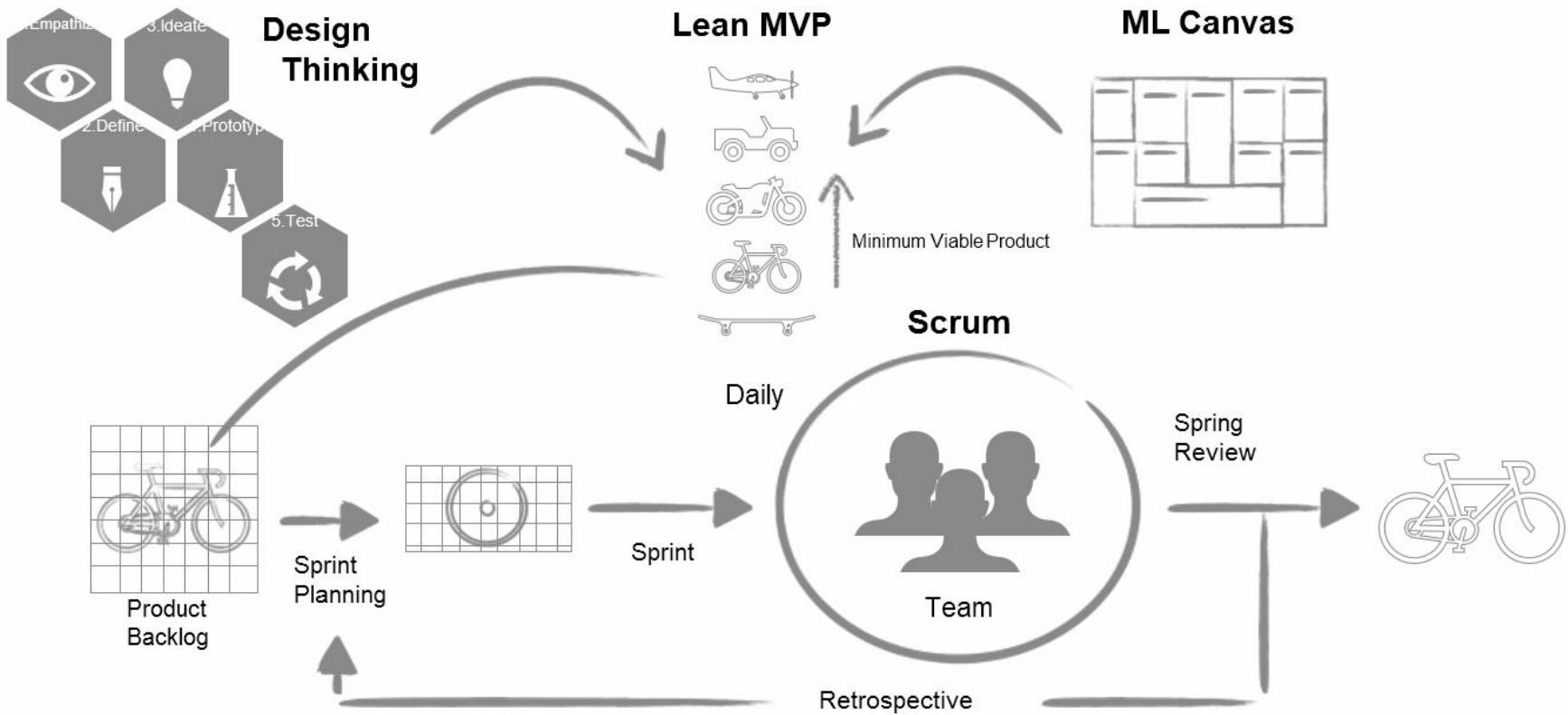
- **¿Qué queremos mirar de manera permanente para saber si el valor persiste?**
- **¿Cómo vamos a hacer las mediciones?**
- **¿Vamos a definir métricas o vamos a usar métricas existentes?**
- **¿El modelo se degrada?**

 OBJETIVOS DE LA ORGANIZACIÓN	 OBJETIVOS DEL PRODUCTO	 DESAFÍO QUE BUSCAMOS RESOLVER	 STAKEHOLDERS	 IMPACTO EN EL USUARIO FINAL
<p>Lograr rentabilidad Lograr fidelidad de nuestros asociados Conseguir crecimiento en el mercado Ofrecer un servicio de calidad</p>	<p>Reducir la tasa de abandono entre los asociados. Mejorar la tasa de éxito de los esfuerzos de retención.</p>	<p>Contexto: Somos una empresa que ofrece prestaciones médicas. Buscamos identificar clientes que pueden llegar a cancelar el servicio para que se tomen las medidas apropiadas.</p>	<p>Área de gestión de relaciones con el cliente (CRM) IT Negocios Marketing</p>	<p>Reducir el tiempo de análisis y permitir que los esfuerzos se centren en determinar la mejor manera de llegar al cliente y así evitar que cancele el servicio</p>
 DATA INPUT Uso de los servicios Consultas realizadas con Atención al cliente Edad Herramientas CRM Base de datos de pago	 OUTPUT DEL SISTEMA Número de 0 al 100 con la probabilidad de que un asociado cancele el servicio 0-40 poco probable / 41 a 70 probable / 71 a 100 crítico			
 COMPETENCIA Actualmente tenemos un servicio de contact center para tratar de gestionar la retención de los asociados. El modelo ayudaría a reducir los costos automatizando las gestiones y evitaría en una gran medida que lleguemos a la instancia de cancelación, ya que la fidelización comenzaría cuando empiezemos a detectar patrones de comportamiento en los asociados.	 RESTRICCIONES ESTRUCTURALES Capacitar al equipo de CRM para la correcta utilización del modelo y la interpretación de sus resultados. Costos para tener un servidor dedicado para recibir la información y correr el modelo		 IMPACTO DEL NEGOCIO Esperamos reducir el número de cancelaciones en un 40%	 CRECIMIENTO Automatizar el modelo para que se entrene todos los meses y así poder obtener predicciones y realizar acciones de marketing automáticas sobre los usuarios con mayor probabilidad de cancelar el servicio
	 EVALUACIÓN Y MONITOREO EN TIEMPO REAL Medir la precisión de las predicciones del mes pasado sobre el conjunto de retención. Evaluar los resultados de la inversión.			

Metodologías de trabajo y buenas prácticas en Data Science



Metodologías que coexisten



El primer paso, definir los objetivos de negocio.

A menos que sepamos qué preguntas hay que responder y qué problema empresarial se desea resolver, no se debe iniciar un proyecto de datos.

- Los gestores deben tener claros los objetivos.
- Los técnicos, entendiendo esto, podrán saber qué datos recopilar y cómo hacerlo.
- De esta manera, evitamos hacer excesos de trabajo y sacar conclusiones erróneas, respuestas o indicaciones que no necesitamos.
- ¡Definir KPIs de negocio es clave!**

Diferenciar entre lo que tenemos y lo que necesitamos

Disponer de una enorme cantidad de datos almacenados no implica, necesariamente, que dispongamos de buenos datos.

- Especificar los datos que queremos almacenar y aquellos de los cuales podemos prescindir.
- Recopilar datos con rigor y procesos definidos de antemano.
- ¡De no hacerlo así, puede que después sea muy tarde y no se pueda volver atrás!**

Formación continua y personal calificado son clave

Si queremos tener más probabilidades de éxito en los proyectos de datos en las empresas, necesitamos no solo personal calificado, sino formación continua.

¡El mundo de la ciencia de datos está en permanente movimiento!

Reflexiones Finales



1. **No hace falta ser un unicornio tecnológico** para empezar a usar la Ciencia de Datos.
2. **No importa si tu empresa no es de base tecnológica**, la Ciencia de Datos es una herramienta transversal a todas las industrias.
3. La Ciencia de Datos no es sólo aplicar algoritmos de Machine Learning. A veces, **sólo con visualizar los datos de una manera inteligente** podemos mejorar enormemente la toma de decisión.
4. El **nuevo paradigma de los negocios** gira en base a los datos y a la interdisciplina.

ACADEMICA