

## PLAN DE IMPLEMENTACIÓN DE LA ESTRATEGIA DATA DRIVEN

### Paso 1: Establece una Estrategia de Datos

- Organizar y descubrir
- Evaluación de los activos digitales
- Análisis, priorización y roadmap
- Gestión del cambio

### Paso 2: Democratiza tus Datos

- Volumen. Se refiere a la enorme cantidad de información producida por los sistemas digitales.
- Velocidad. Se refiere al ritmo al que se crean, ingieren y procesan los datos.
- Variedad. Se refiere a los formatos y fuentes de datos heterogéneos que se deben cruzar en procesos de análisis.
- Variabilidad. Describe la evolución de los datos ingeridos.
- Veracidad. Describe el grado en que los datos son exactos, precisos y fiables.

### Paso 3: Construye una Cultura de Recogida de Datos

- a) Mediciones directas
- b) Observación activa y pasiva
- c) Estudios de casos
- d) Encuestas

### Paso 4: Acelera la velocidad de comprensión

La idea detrás de este principio es democratizar la información y su comprensión dentro de la empresa. Si proporcionas una alta velocidad y comprensión dinámica a las tomas de decisiones, estas se convertirán en decisiones fiables. Es este sentido, una organización basada en los datos toma todas y cada una de sus decisiones basadas en ellos.

Para ello, es importante que uso de estos Datos genere la mayor comprensión posible. Una de las mejores maneras de desencadenar esta comprensión dentro la empresa es utilizar herramientas de gráficos dinámicos. Estas herramientas proporcionan una mejor visión de los Datos a la hora de la toma de decisiones.

### Paso 5: Mide el valor del Data Science

- **Regresión lineal:** La cuna del Machine Learning. **Un algoritmo básico que se aprende ya en el instituto. Introduce conceptos como la beta de las variables, p-valor o R Cuadrado.** Todo científico de datos debería conocer estos conceptos y saber interpretarlos a la perfección.
- **Regresión logística:** Una modificación de la regresión lineal para poder resolver problemas de clasificación. **Aproximadamente el 80% de los proyectos de Machine Learning son problemas de clasificación.** Aquí se aprenden conceptos como la diferencia entre efecto multiplicativo y aditivo.
- **Series temporales:** Como decía Lewis Carroll en Alicia en el país de las maravillas “El tiempo es todo un personaje”. Y para estudiarlo, es necesario conocer las series

temporales. **Otra pequeña modificación de las regresiones lineales, donde se introducen conceptos como estacionalidad y tendencias** a través de los modelos ARIMA.

- **Árbol de decisión:** Quizás el más entendible de los algoritmos, de cara a presentar resultados. **Los árboles con sus hojas y sus ramas han dado pie a los algoritmos más usados en la actualidad, Random Forest** (no se hace un árbol, se hacen miles y se ve lo que se aprende de todos ellos) y **XGBoost** (cada vez que se hace un árbol, se observa donde se equivoca y se corrige con otro árbol).
- **Red Neuronal:** Uno de los algoritmos más potentes y más oscuros, **sabemos que acierta pero no el porqué**. Cada vez se están sacando mas herramientas para intentar que esa oscuridad se convierta en luz y entender cómo aciertan las redes. **Basadas en el cerebro humano, es básico conocer conceptos como capa, neurona o retropropagación**. Las técnicas que usa Deep Learning, ahora tan de moda, sobre todo en el uso de la inteligencia artificial para detección de imágenes, voz o comprensión de texto, **están basadas en redes neuronales**. Estar actualizados en esta y otras técnicas es imprescindible para no perder competitividad.
- **K-means:** Otra de los algoritmos más conocidos, en este caso para hacer clustering. **Nos sirve para agrupar y clasificar por características similares**. Conceptos como normalización o distancias son importantes en este algoritmo.
- **k-NN:** También llamado k-vecinos. **Usado para los recomendadores. ¿Qué te voy a recomendar? Lo mismo que a tus vecinos**, entendiendo como vecinos las personas que tengan tus mismos hábitos.
- **PCA:** Análisis de componentes principales, es **una de las técnicas llamadas de reducción de dimensionalidad**. Muy usada en las encuestas. ¿Por qué usar cientos de variables si podemos utilizar solo unas pocas que resuman la información? **Conceptos como rotación son interesantes de aprender en estas técnicas**.
- **Cadenas de Markov:** Sirven para **introducir la estadística bayesiana y se ven conceptos como simulación**. Muy útil para cuando tenemos pocos datos, pues con estas técnicas se pueden dar pistas al modelo para que aprenda.
- **Simplex :** Este algoritmo está **centrado en buscar soluciones óptimas con restricciones**. Es **el último paso de los modelos, donde puedes pasar de predecir a ser prescriptivo**. Predecir es ver las posibles ventas de una empresa bajo ciertas hipótesis. Ser prescriptivo es decir qué palancas hay que tocar para maximizar las ventas. Es **el último paso para sacar todo el rendimiento a los algoritmos**, y el método del simplex es la base para entender cómo se puede realizar esto.

## **Paso 6: Aplica una estructura de gobernanza de datos**

### Director de datos / Comité directivo (Nivel Estratégico)

- Supervisión del programa de Gobierno de Datos.
- Aseguramiento, aprobación y la dotación de personal para el programa.
- Lidera y vigila el el programa en su constante en su progreso y actuar como defensor del mismo internamente.
- Define las directrices para crear y mantener los datos.

Comité de Gobierno de Datos / Junta de Gobierno de Datos (Nivel Táctico)

- Ejecución del programa de Gobierno de Datos.
- Definición de las reglas, procesos y políticas del programa.
- Evalúa los requisitos y problemas con datos.
- Mide e informa de la calidad de los datos y documentación de los mismos.

Administrador de Datos / Arquitecto de datos (Nivel Operativo)

- Supervisión de los conjuntos de datos.
- Perfil de definiciones y formatos de elementos de datos estandarizados.
- Garantizar las políticas, reglas y asignaciones aprobadas en los niveles superiores.
- Evaluación de requisitos, problemas con los datos y apoyo de proyectos

Propietarios de Datos (Nivel Operativo)

- Especifica los requisitos empresariales en los datos y su calidad.
- Participación activa en la creación / revisión / aprobación de definiciones de datos.
- Integridad y calidad de la definición de datos.

Operador de Datos (Nivel Operativo)

- Crear, actualizar, eliminar y mantener datos basados en los estándares definidos.
- Se encargan de la operación del ciclo de vida de los datos en los parámetros definidos
- Integridad del uso de datos.
- Identificación y documentación de problemas de riesgo, normativos y/o legales.

