



# LLEGANDO AL CONOCIMIENTO Y A LA SABIDURÍA

AUTORA: JIMENA DEL CAMPO



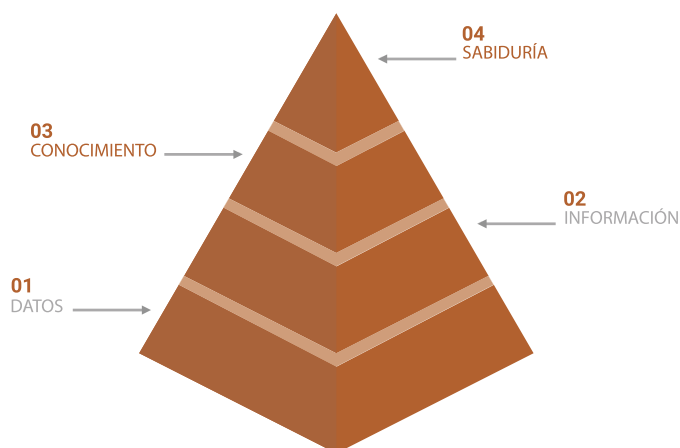


# CONTENIDO

INTRODUCCIÓN .....	3
1. ANÁLISIS PREDICTIVO .....	4
2. CUÁNDO APLICAR MACHINE LEARNING .....	7
3. TRANSFORMANDO OBSERVACIONES EN PREDICCIONES.....	10
4. PASOS PARA IMPLEMENTAR UN MODELO DE MACHINE LEARNING .....	15
5. ANALÍTICA PRESCRIPTIVA .....	17
BIBLIOGRAFÍA.....	19

# INTRODUCCIÓN

Hasta ahora hemos podido depurar los datos, darles un contexto para transformarlos en información y así visualizarlos para avanzar con el análisis de hechos históricos.



Pero ¿cómo hacemos para avanzar más allá y hacer un análisis del futuro, para poder predecir lo que sucederá en nuestros negocios, así como implementar la mejor opción de forma automática?

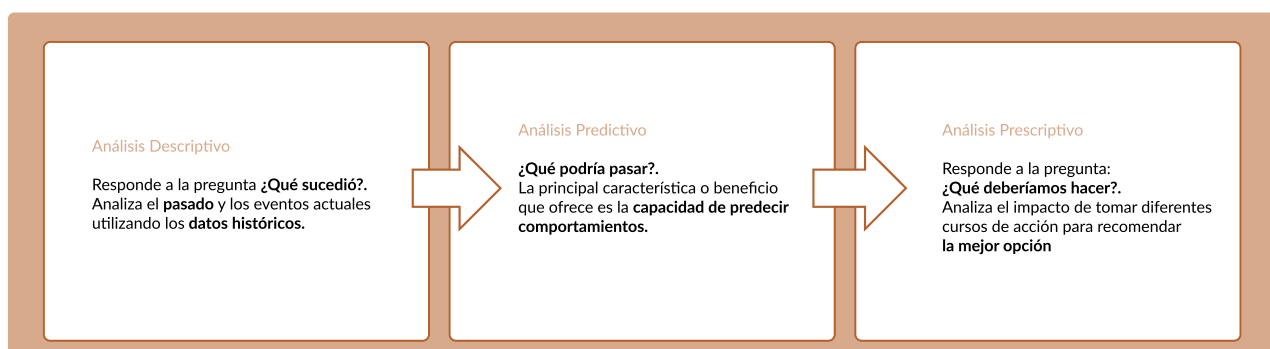
En esta unidad nos adentraremos en la analítica predictiva y prescriptiva que nos permitirán mejorar nuestras decisiones, llegando a través de ellas a la cúspide de la pirámide del conocimiento.





# ANÁLISIS PREDICTIVO

El análisis predictivo es el que le sucede al descriptivo, ya que basado en datos históricos tiene la capacidad de predecir comportamientos futuros.

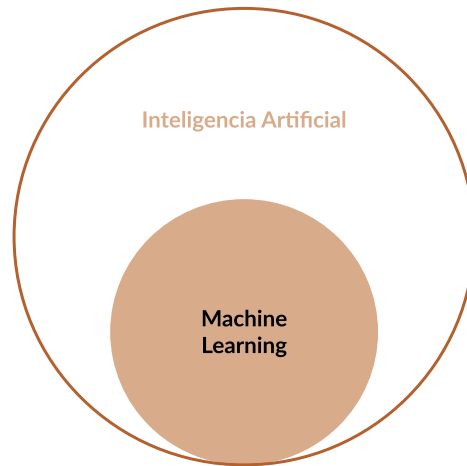


Es decir que el análisis descriptivo mira el ayer y el ahora, el análisis predictivo mira el mañana y el análisis prescriptivo permite, a futuro, implementar la mejor de las decisiones.

Avancemos ahora con el análisis predictivo. ¿Cómo podemos hacer para predecir basados en datos históricos? Y es aquí donde aparece el concepto de Machine Learning. Veamos a continuación algunas definiciones:



**Inteligencia Artificial (también conocida como IA):** es la inteligencia demostrada por las máquinas, en contraste con la inteligencia natural de los humanos. Esta tecnología permite programar una computadora para que haga algo, indicando a la máquina lo que debe hacer. Es decir que la máquina no aprende por su cuenta y solo actualizará su forma de accionar cuando explícitamente se la programe para ello, a través de enseñarle o de programar nuevas reglas.



**Machine Learning (Aprendizaje Automático):** podríamos decir que son **algoritmos** que se encargan de analizar datos históricos y detectar patrones que un humano no podría detectar, para luego hacer predicciones. En este caso se le proporcionan datos a una computadora y esta aprende por sí sola, mejorando las tareas que realiza a medida que adquiere experiencia. Es un subconjunto de la **Inteligencia Artificial**.

**Algoritmo:** es una secuencia de pasos lógicos que permiten la solución de uno o más problemas.



Aclaremos un punto con respecto al Machine Learning: cuando decimos que la computadora “aprende por sí sola” nos referimos a la capacidad del algoritmo de detectar patrones automáticamente sobre los datos históricos.



Por otro lado, cuando este algoritmo “aprendió” de los datos históricos, también se suele decir que el algoritmo está “entrenado”, transformándose en un programa ejecutable capaz de tomar datos históricos y hacer predicciones con cierta frecuencia. A esto, más todo el preprocesamiento previo de datos que serán input del algoritmo, se lo conoce como **Modelo de Machine Learning**.

Ahora bien, un concepto que se encuentra totalmente relacionado con Machine Learning es el **Big Data**, ya que este último implica cantidades de datos que son muy grandes y que se generan a gran velocidad y de diversas fuentes, por lo que requiere de prácticas y tecnologías avanzadas para su procesamiento. Pero, a pesar de estas grandes cantidades de datos, los algoritmos de Machine Learning suelen ser muy veloces, llegando a procesar millones de datos por segundo.

La generación de datos sigue creciendo a pasos agigantados y la evolución de la tecnología también. Esto nos puede llevar a pensar a que si generamos tantos datos entonces todas las empresas deberían estar utilizando *Machine Learning* como apoyo a sus decisiones, no obstante, la verdad es que, si no se tiene la madurez analítica suficiente y una base sólida de data, no se podrán sortear fácilmente los retos del uso del Machine Learning. Casi el 90% de los proyectos de ML nunca llegan a cumplir su objetivo final.



# CUÁNDO APLICAR MACHINE LEARNING

Muchas empresas quieren ser pioneras en tecnologías y comienzan por implementar Analítica Predictiva en sus negocios, sin evaluar previamente la madurez de los datos, del negocio y de la compañía. Como consecuencia, terminan gastando grandes sumas de dinero en proyectos que no se implementan.

Tal fue el caso de Netflix en el año 2009 que a través de un concurso premió con un millón de dólares a un equipo que mejoró el Sistema de Recomendación que estaban usando y si bien la solución cumplió con el objetivo perseguido, el esfuerzo para su implementación era tan grande que no valió la pena ponerlo productivo.

Pero veamos ahora algunos ejemplos de negocios en los que podemos implementar *Machine Learning*.

- Predicción de la cantidad de unidades a vender de un producto en particular.
- Predecir la probabilidad de compra de un producto.
- Predecir un mal comportamiento de pago.
- Predecir la posible baja o abandono de un cliente: es decir si el cliente va a dejar de comprar en la empresa.
- Predecir posibles transacciones fraudulentas.



También en nuestro día a día encontramos casos sencillos en los que implementamos Machine Learning:

- Clasificación de emails como *Spam*.
- Recomendaciones de música en Spotify.
- Sistemas de recomendación de las plataformas de *Streaming* (Netflix / Disney Plus, etc.), que se basan en las visualizaciones que realizamos y las que realizan los demás usuarios para sugerirnos los productos que podemos consumir.
- Sistemas de recomendación de plataformas de *e-commerce* (Mercado Libre / Amazon).
- Waze, cuando nos indica elementos valiosos al momento de conducir, tales como la presencia de accidentes, objetos peligrosos en la carretera, radares o agentes de tráfico y que, al igual que *Google Maps*, también nos indica la mejor ruta.

Es muy común que nos encontremos con empresas que desean llegar al Machine Learning sin haber pasado antes por la Analítica Descriptiva, fracasando en sus intentos de poner a funcionar tecnología para la cual aún no están preparados. Muchas veces desde la analítica descriptiva podemos encontrar patrones de comportamiento y deducir reglas sencillas que nos permitan lograr grandes resultados de forma rápida y a bajo costo, teniendo en cuenta el conocimiento del negocio.

Todas estas herramientas tienen los siguientes puntos en común:

- Facilidad de uso
- Múltiples fuentes de datos a las cuales conectarse
- Variedad de tipos de tablas, gráficos y mapas
- Documentación, tutoriales y soporte diseñados de manera intuitiva para el usuario
- Pueden manejar grandes conjuntos de datos e incluso múltiples conjuntos de datos en una sola visualización
- Permiten compartir información
- Delivery de informes vía email
- Exportación de información a formatos conocidos (excel, CSV, Spreadsheet, PDF, imagen)





- Todas ofrecen un período de prueba sin costo, que permite conocer más acerca del uso de cada una, a excepción de DataStudio de Google que es totalmente gratis si tienes cuenta de Google.

En este ejemplo vamos a ver cómo se construye el siguiente informe en Data Studio a partir de una hoja de cálculo de Google y aprenderemos más acerca de los conceptos de visualización:



# TRANSFORMANDO OBSERVACIONES EN PREDICCIONES

Comprendamos un poco más a fondo cómo funciona *Machine Learning*. Cuando comenzamos un proyecto de *Machine Learning* es necesario comprender a la perfección cuál es el objetivo que perseguimos, es decir qué resultado estamos esperando de la ejecución del algoritmo.

Analicemos la siguiente situación: dado que cada vez que voy al trabajo y llego a tiempo, recibo una recompensa monetaria en mi bono mensual, me urge saber con precisión qué debo hacer para llegar a tiempo al trabajo todos los días, teniendo en cuenta que los días lunes y viernes debo utilizar la moto para viajar, mientras que de martes a jueves puedo utilizar el auto o carro. Mi objetivo es poder determinar si llego a tiempo o no al trabajo.

Por otro lado, debo identificar cuáles son todos los factores o variables que pueden influir en el resultado perseguido, tales como: tipo de vehículo, hora de salida, día de la semana, estado del tiempo, estado de la ruta, estado del tránsito, etc.

Finalmente, es necesario que vaya registrando todos los días el resultado u observaciones que van ocurriendo.



## Pasos a seguir en un proyecto de *Machine Learning*:

- Identificar la variable objetivo
- Identificar las variables o factores que influyen en el objetivo o resultado buscado
- Registrar los resultados u observaciones obtenidas



Para este fin, podríamos preparar una tabla como la siguiente:

	A	B	C	D	E	F	G	H
1	Observaciones	Objetivo	Tipo de vehículo	Hora de salida	Día de la semana	Estado del tiempo	Estado de la ruta	Estado del tránsito
2								
3								
4								
5								
6								
7								

En la columna **observaciones** iré anotando el día en el que se produce la observación.

En la columna **objetivo**, iré anotando el resultado de la observación, es decir si llegué o no a tiempo.

A partir de la tercera columna, vamos a ir registrando el valor de las distintas variables que pueden incidir en nuestro resultado. En la columna **tipo de vehículo** iré anotando si utilicé auto o moto para trasladarme.

Luego, la **hora a la que salí de casa**, qué **día de la semana** es. También me interesa saber el estado del tiempo en ese día, si estaba despejado, lluvioso, ventoso o nublado.

A su vez, considero importante conocer el **estado de la ruta** para detectar si está en buen estado o si hay algún corte por reparación.

Finalmente, y no menos importante, registrar cómo está el **tránsito** ese día, definiendo si está congestionado o si es fluido.

El día lunes, realizo mi primer registro, anotando los siguientes valores y dejando como resultado que no llegué a tiempo.

Observaciones	Objetivo	Tipo de vehículo	Hora de salida	Día de la semana	Estado del tiempo	Estado de la ruta	Estado del tránsito
Día 1	No a tiempo	Moto	6:45 am	Lunes	Despejado	Buen estado	Congestionado

De aquí puedo deducir que no tuve mucha suerte con el tránsito.



Al día siguiente me voy en auto y salgo de casa a las 7 am, registrando los siguientes valores:

Observaciones	Objetivo	Tipo de vehículo	Hora de salida	Día de la semana	Estado del tiempo	Estado de la ruta	Estado del tránsito
Día 1	No a tiempo	Moto	6:45 am	Lunes	Despejado	Buen estado	Congestionado
Día 2	No a tiempo	Auto	7 am	Martes	Despejado	Buen estado	Fluido

Comparando estos valores con los del primer registro, vemos que cambió el tipo de vehículo, el horario de salida, el día de la semana y me acompañó el estado del tránsito.

Si bien todas las condiciones estaban dadas para que llegara a tiempo, puedo deducir que no fue una buena idea salir a las 7 am, en auto, el día martes.

Por lo que el día miércoles decido salir más temprano y si bien el clima no me acompaña, logro llegar a tiempo al trabajo.

Observaciones	Objetivo	Tipo de vehículo	Hora de salida	Día de la semana	Estado del tiempo	Estado de la ruta	Estado del tránsito
Día 1	No a tiempo	Moto	6:45 am	Lunes	Despejado	Buen estado	Congestionado
Día 2	No a tiempo	Auto	7 am	Martes	Despejado	Buen estado	Fluido
Día 3	A tiempo	Auto	6:45 am	Miércoles	Lluvioso	Buen estado	Fluido

Esta es la primera vez que logro mi objetivo y si comparo los valores con los del primer y segundo registro, puedo deducir que el cambio de horario fue favorable.

Lo que me interesa saber en este punto es cómo las columnas de la derecha, desde tipo de vehículo hasta estado del tránsito, están afectando al objetivo que por supuesto es: llegar a tiempo al trabajo todos los días.

A medida que tenemos mayor cantidad de observaciones, gran parte de nuestro aprendizaje como humanos empieza a generarse de esta forma: diferentes variables con valores diferentes van arrojando resultados diferentes y poco a poco podemos deducir qué día de la semana es más probable que lleguemos a tiempo, cuál es el tipo de vehículo que nos conviene utilizar, a qué hora debo salir de casa, etc.

Ahora bien, ¿qué cantidad de registros podemos tomar en nuestra vida como muestra? Si consideramos que cada año tiene 365 o 366 días, dependiendo de si es bisiesto o no, y descontamos los fines de semana (52 sábados y 52 domingos), podríamos decir que aproximadamente podríamos registrar entre 262 y 261 observaciones en un año. ¿Qué hacemos entonces cuando el número de observaciones crece tanto?





Seguramente a medida que el tiempo va pasando vamos teniendo mayor expertise y logramos llegar a tiempo al trabajo con mayor frecuencia. A medida que tenemos más experiencia aprendemos más.

Observaciones	Objetivo	Tipo de vehículo	Hora de salida	Día de la semana	Estado del tiempo	Estado de la ruta	Estado del tránsito
Día 1	No a tiempo	Moto	6:45 am	Lunes	Despejado	Buen estado	Congestionado
Día 2	No a tiempo	Auto	7 am	Martes	Despejado	Buen estado	Fluido
Día 3	A tiempo	Auto	6:45 am	Miércoles	Lluvioso	Buen estado	Fluido
Día 4	No a tiempo	Auto	6:45 am	Jueves	Nublado	Buen estado	Congestionado
Día 5	A tiempo	Moto	6:45 am	Viernes	Nublado	Buen estado	Fluido
Día 6	No a tiempo	Moto	6:45 am	Lunes	Nublado	Buen estado	Congestionado
Día 7	No a tiempo	Auto	6:45 am	Martes	Despejado	Corte por reparación	Fluido
Día 8	No a tiempo	Auto	6:55 am	Miércoles	Nublado	Buen estado	Fluido
Día 9	A tiempo	Auto	7 am	Jueves	Nublado	Buen estado	Congestionado
Día 10	No a tiempo	Moto	6:45 am	Viernes	Nublado	Buen estado	Fluido
....	....	....	....	....	....	....	....
Día 262	No a tiempo	Moto	6:45 am	Viernes	Nublado	Buen estado	Fluido

Otro factor que tenemos en cuenta es el número de variables que tenemos a la derecha, es decir aquellas que influyen en el resultado. Seguramente, a medida que vamos teniendo mayor experiencia comenzamos a analizar otros factores, tales como si hubo o no un accidente de tránsito, si es un día festivo, el estado del vehículo, si tuve que ir a cargar gasolina, etc.

Observaciones	Objetivo	Tipo de vehículo	Hora de salida	Día de la semana	Estado del tiempo	Estado de la ruta	Estado del tránsito	Accidente de tránsito	Día festivo	Variable N
Día 1	No a tiempo	Moto	6:45 am	Lunes	Despejado	Buen estado	Congestionado	no	no	A
Día 2	No a tiempo	Auto	7 am	Martes	Despejado	Buen estado	Fluido	si	no	B
Día 3	A tiempo	Auto	6:45 am	Miércoles	Lluvioso	Buen estado	Fluido	no	no	A
Día 4	No a tiempo	Auto	6:45 am	Jueves	Nublado	Buen estado	Congestionado	no	no	C
Día 5	A tiempo	Moto	6:45 am	Viernes	Nublado	Buen estado	Fluido	no	no	B
Día 6	No a tiempo	Moto	6:45 am	Lunes	Nublado	Buen estado	Congestionado	no	no	A
Día 7	No a tiempo	Auto	6:45 am	Martes	Despejado	Corte por reparación	Fluido	no	no	B
Día 8	No a tiempo	Auto	6:55 am	Miércoles	Nublado	Buen estado	Fluido	no	no	C
Día 9	A tiempo	Auto	7 am	Jueves	Nublado	Buen estado	Congestionado	no	si	C
Día 10	No a tiempo	Moto	6:45 am	Viernes	Nublado	Buen estado	Fluido	no	no	C
....	....	....	....	....	....	....	....	....	....	....
Día 262	No a tiempo	Moto	6:45 am	Viernes	Nublado	Buen estado	Fluido	no	no	A



¿Cuántas variables creen que podemos tener en cuenta dentro de nuestro análisis? Podrán ser tantas como nuestra capacidad cognitiva lo permita.

Si observamos, a medida que pasa el tiempo nuestra experiencia crece, lo que nos permite registrar mayor cantidad de observaciones y mayor cantidad de variables.

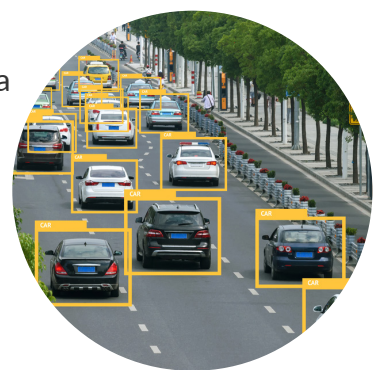
Pero el esquema en general sigue siendo el mismo: Cómo influyen las variables de la derecha para lograr predecir el objetivo que nos propusimos, mientras registramos las distintas observaciones:

Observaciones	Objetivo	Variable 1	Variable 2	Variable 3	Variable 4	...	Variable n

Entonces ¿cómo hacemos para evaluar toda la información que vamos recabando? Les puedo asegurar que después de un año de observaciones y con más de 10 variables, se torna muy difícil para una persona poder evaluar todas las posibles situaciones que se pueden dar.

Adicionalmente, si a este registro le agregamos las observaciones de todas las personas que comparten mi misma ruta, donde el nro. de observaciones ya no será sólo de 261, sino que se irá multiplicando por la cantidad de experiencias que podemos registrar, entonces probablemente ya no podamos seguir sacando conclusión alguna.

Y es en este punto donde empezamos a hablar de BigData, cuando la cantidad de información con la que contamos crece y necesitamos de otras herramientas de procesamiento, tales como el Machine Learning. Donde la máquina va barriendo las distintas observaciones y variables y va aprendiendo (encontrando nuevos patrones que el cerebro humano no podría encontrar con la agilidad que lo hace un modelo de *Machine Learning*).



Si bien el proceso de definición del modelo es sencillo de comprender, debemos saber que los proyectos de *Machine Learning* no son sencillos de implementar si no contamos con el equipo adecuado de Científicos de Datos que sean capaces de evaluar cuál es el algoritmo que mejor se adapta a la necesidad u objetivo que estamos persiguiendo para luego implementar el modelo completo que se basa en los conceptos mencionados anteriormente, pero que requiere de la ejecución de los siguientes pasos para su ejecución:

# PASOS PARA IMPLEMENTAR UN MODELO DE MACHINE LEARNING

- a) **Recolectar los datos:** si bien ya hemos hablado sobre este punto y las diversas formas de recolección de datos que existen en el mercado, vale aclarar que este paso es uno de los que más complicaciones trae y más tiempo consume.
- b) **Preprocesar los datos:** una vez que tenemos los datos, tenemos que asegurarnos que tiene el formato correcto para nutrir nuestro algoritmo de aprendizaje. Es inevitable tener que realizar varias tareas de preprocesamiento antes de poder utilizar los datos.
- c) **Explorar los datos:** una vez que ya tenemos los datos y están con el formato correcto, podemos realizar un pre análisis para corregir los casos de valores faltantes o intentar encontrar a simple vista algún patrón en los mismos que nos facilite la construcción del modelo. En este punto podemos detectar valores atípicos que debemos descartar, o encontrar las características que más influencia tienen para realizar una predicción.
- d) **Entrenar el algoritmo:** aquí es donde comenzamos a utilizar las técnicas de Machine Learning realmente. En esta etapa nutrimos al o los algoritmos de aprendizaje con los datos que venimos procesando en las etapas anteriores.
- e) **Evaluar el algoritmo:** en esta etapa ponemos a prueba la información o conocimiento que el algoritmo obtuvo del entrenamiento del paso anterior. Vemos que tan preciso es el algoritmo en sus predicciones y si no estamos muy conforme con su rendimiento, podemos volver a la etapa anterior y continuar entrenando el algoritmo cambiando algunos parámetros hasta lograr un rendimiento aceptable o bien seleccionar otro algoritmo que se adapte mejor a nuestra necesidad.



**f) Utilizar el modelo:** es la última etapa, donde ponemos a nuestro modelo a enfrentarse al problema real.

Podemos decir que el fundamento de Machine Learning es traducir características de un conjunto de observaciones a predicciones y si bien existen varios esquemas de representación y según el tipo de variable distintas clasificaciones (por ej: clasificación y regresión), es importante que entendamos la forma en la que se procesa la información y porqué llega un punto en el que los humanos requerimos de las máquinas para procesar cantidades tan grandes de información.



Recuerden siempre comenzar por la solución más sencilla, aunque eso implique no usar Machine Learning. Solo debemos utilizar Machine Learning si nuestra habilidad cognitiva no puede por sí sola, logrando obtener un beneficio mayor al costo de su implementación.





# ANALÍTICA PRESCRIPTIVA

Hasta ahora hemos recorrido la pirámide logrando llegar al conocimiento y comprendiendo cómo funciona la analítica predictiva, no obstante, ¿cómo damos el paso hacia la analítica prescriptiva?

Recordemos la definición de Analítica Prescriptiva: el Análisis Prescriptivo responde a la pregunta: ¿qué deberíamos hacer? Analiza el impacto de tomar diferentes cursos de acción para recomendar la mejor opción.

A diferencia del análisis descriptivo o el predictivo, el prescriptivo determina los distintos modos en que los procesos de negocios deben evolucionar o ser modificados.

La analítica prescriptiva utiliza una combinación de técnicas y herramientas tales como reglas de negocio, algoritmos de aprendizaje automático y procedimientos de modelado computacional. Estas técnicas se aplican a los diferentes conjuntos de datos, incluyendo datos históricos y transaccionales, se alimenta de datos en tiempo real y en general de grandes volúmenes de datos.

Por último, podemos decir que la analítica prescriptiva se basa en el conocimiento y las técnicas de las analíticas descriptiva y predictiva, adicionando la investigación operativa y la optimización numérica.

Cada negocio es distinto, por lo que no existen técnicas de analítica prescriptivas genéricas, sino que todo proyecto comienza con un trabajo de consultoría para terminar en un proyecto hecho a medida de las necesidades del negocio que lo requiere.



Veamos algunos casos de aplicación:

- Una empresa de Internet se da cuenta de que el uso que hace un cliente de sus servicios está disminuyendo. El análisis prescriptivo puede sugerir alguna optimización de los servicios o bien un ajuste en los precios para evitar perder a ese cliente.
- Una empresa que posee servicios de logística donde según la demanda futura puede definir qué mercadería dejar acopiada en los distintos almacenes, optimizando de esta forma el almacenamiento y el delivery de sus productos.
- Realizar un mantenimiento predictivo en aviones, sustituyendo las revisiones periódicas o las reparaciones en caso de fallo por un sistema que recomiende el mejor momento para realizar la revisión de cada componente, reduciendo costes por averías y reparaciones.



# BIBLIOGRAFÍA

JOHNSTON, C. (2012) Caso Netflix - Sistema de recomendación no implementado. Recuperado de: <https://www.wired.com/2012/04/netflix-prize-costs/>

THEOBALD, O. (2017) Machine Learning for Absolute Beginners: A Plain English Introduction (Second Edition).

