

Analytics y Big Data: qué son y cómo se aplican en la industria del petróleo y del gas

Por *Ing. Estanislao Martín Irigoyen* (Intelligentia S.R.L.)

Se conoce como *Big Data* al conjunto de tecnologías para trabajar con grandes volúmenes de datos que, debido a su heterogeneidad, tienen grandes variabilidades y necesitan una gran velocidad de procesamiento para sustentar soluciones operables.

Hablar de Minería de Datos (*Data Mining*), *Analytics*, *Big Data Analytics* o *Fast Data Analytics* en la actualidad se ha tornado moneda corriente. Con mayor frecuencia, día a día los medios masivos de comunicación colocan alguna referencia a estas nuevas (o no tanto) disciplinas con fuerte base en tecnologías de la información y con un gran impacto en el negocio y la resolución de sus problemas más relevantes.

Si bien las industrias del Petróleo y Gas o de la Energía no son ajenas a estas tendencias, los hechos muestran que son las últimas en sumarse a esta gran ola de innovación. Una “ola” capaz de dotar al negocio de una capacidad transformadora que puede volver viables determinados procesos, que por sus costos, hubiesen resultado de aplicación impensable en épocas de precios bajos como las actuales.

Centradas en el rol de ser una gran ventaja competitiva y con muchos éxitos cosechados a lo largo del tiempo en otras industrias, la Minería de Datos o un concepto más generalista, el *Analytics*, se presenta a la industria del Petróleo, del Gas y de la Energía para aportar soluciones a los más variados problemas ayudando a optimizar procesos, minimizar incertidumbres y mejorar la toma de decisiones de los ejecutivos, geólogos, geofísicos e ingenieros de la industria.

Big Data, *Fast Data*, *Real Time* o CEP son términos asociados y de uso frecuente en la actualidad. Los mismos se encuentran ligados al *Analytics*, ya que es esencialmente esta disciplina la que realiza un uso intensivo de las características que las primeras aportan al negocio: El almacenamiento, la gestión, el procesamiento de volúmenes “monstruosos” de datos y/o el accionamiento operativo y en tiempo real (es decir ¡a la velocidad a la que el dato se produce!) se verifique en la medida que los eventos se suceden en el tiempo y según sea el caso.

Esta nota pretende abordar dichos conceptos con un conjunto de ideas y conceptos que permitan al lector un acercamiento amigable a la temática.

¿De qué hablamos cuando hablamos de *Analytics*?

Hablar de *Analytics* no es tarea fácil, ya que puede quedar muy “soft” o bien muy “techie” según quien sea el lector. Lo cierto es que *Analytics* es una disciplina con un altísimo grado de multidisciplinariedad, que involucra tanto a los sectores técnicos como a los sectores del negocio en un mismo proceso. Por lo tanto, se darán un conjunto de definiciones que aporten a todos los lectores, sin pecar en generalidades ni en especificidades.

Analytics: algunas definiciones

“*Analytics* es un conjunto de avanzadas técnicas matemáticas, de programación entera y no lineal, de estadística multivariada, de tecnologías de información, visualización de información e inteligencia artificial que aportan al analista un incremento en su capacidad de buscar y detectar **patrones ocultos** en los datos, para luego exponer el conocimiento extraído de los mismos en modelos lógico-matemáticos que permitan a los equipos de trabajo colaborar en la resolución de problemas”.

Una definición de esta índole puede ser comprendida por un usuario técnico pero lo complejo de su formulación trae aparejado el riesgo de despertar “a un soñador empedernido” que pueda pecar en la fijación de sus expectativas dejándolas por fuera de la realidad y/o de los alcances posibles de la disciplina. Una alternativa con que suelo abordar al tema, reduciendo estos riesgos es con una definición orientada a las soluciones. Con esta pueden comprenderse los alcances de la disciplina y fijarse expectativas más realistas o cercanas a las posibilidades de la disciplina en sí.

“*Analytics* es una disciplina iterativa e interactiva, que ayuda al negocio a explorar sus datos, a encontrar valor oculto en los mismos, dejando expuestos los patrones existentes en los mismos como modelos legibles y conocimiento accionable que ayuden a mejorar la rentabilidad de la empresa, ya sea por una reducción de costos apreciable, por un aumento en la eficacia y/o eficiencia de un proceso y a veces a alcanzar a ambos objetivos de conjunto”.

Si bien esta definición no es específica en cuanto al “cómo y con qué se hacen las cosas”, ofrece claridad en torno a lo que *Analytics* puede aportar, dando una idea de cómo se trabaja en esta disciplina (disciplina iterativa e interactiva) y fundamentalmente explica qué es lo esperable como producto de su implementación.

Patrones y modelos: ¿qué son?

Por “patrón” entendemos a un conjunto de elementos (datos en nuestro caso) que forman en conjunto y/o por partes una unidad diferenciada comportamiento y que se repite sistemáticamente a lo largo del tiempo y/o del espacio (Figura 1).

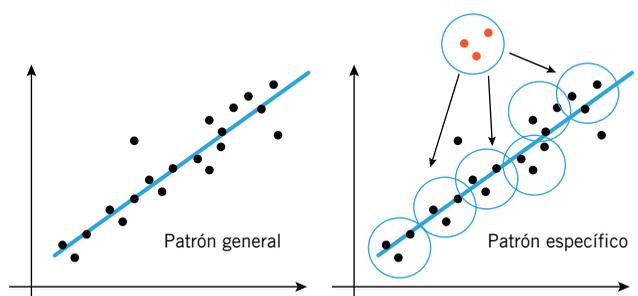


Figura 1.

Los **patrones** pueden ser de tipo correlacional, agrupativo, asociativo, secuencial y eventual, entre otros. Pueden ser patrones en una sola variable (univariados, como el conjunto de cantor) o con la participación de múltiples variables (multivariados). Los Patrones pueden ser también lineales o no lineales (Figura 2).

Por **modelo** entendemos una representación “simplificada” (para algunos) de la realidad, la cual nos ayuda a comprender un conjunto determinado de características de una situación, de un fenómeno, de un comportamiento y/o de un sistema.

Ejemplos de modelos pueden ser la ecuación de una recta de regresión, un conjunto de reglas tipo Si... EN-

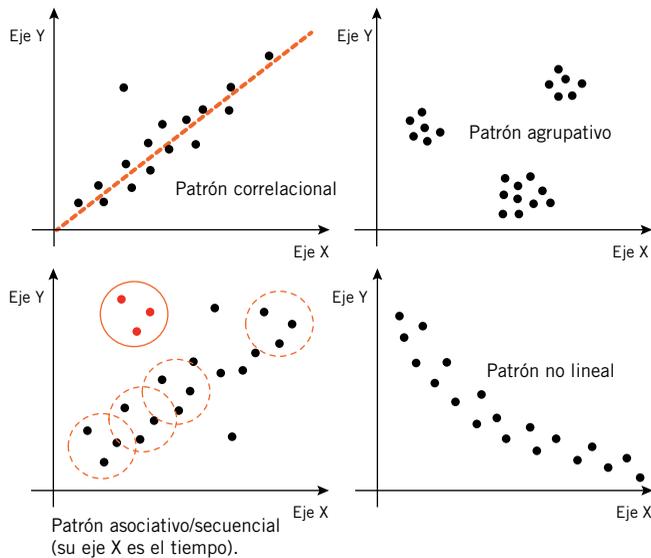


Figura 2.

TONCES..., y otros formatos más complejos, como redes neuronales o conjunto de ecuaciones (reglas y/o fórmulas matemáticas, probabilísticas y/o lógicas).

Será entonces la tarea del **científico de datos** la de adquirir, procesar y modelar características de un sistema (o de un problema en particular de dicho sistema) de manera de comprenderlo y de colaborar con otros equipos técnicos en el diseño de una solución.

Causalidad y correlación, la esencia de todo

Detrás de estas nuevas tecnologías existe un fin último: la obtención de conocimiento accionable (como manifiesta la segunda definición de *analytics*).

Si bien estas herramientas nos ayudan a analizar información y a encontrar patrones en los voluminosos conjuntos de datos de las organizaciones, hasta el momento no pueden reemplazar el poder de razonamiento de un ser humano.

Un algoritmo permite detectar y exponer patrones y/o correlaciones que en esencia se manifiestan como comportamientos conjuntos y/o regulares en una o más variables que miden un proceso o fenómeno. Por otro lado, un algoritmo no puede atribuir estos comportamientos a una causa o concluir que ese comportamiento sea efecto de una causa previa en particular.

Las técnicas de *analytics* permitirán encontrar relaciones entre los datos, sin tener el poder de contextualizar y establecer una relación causal.

Mientras que una relación causal puede utilizarse para realizar predicciones y deducciones, una relación correlacional no. Si existe causalidad, habrá correlación pero no a la inversa, pero es posible la existencia de una relación correlacional, producto de una casualidad pura o de un comportamiento complejo e indirecto, que puede cambiar en cualquier momento. Mientras que una relación causal permite establecer una generalización, una relación correlacional no lo permite y un cambio en una condición (endógena o exógena) puede romper esa relación.

Este punto es quizá el más importante de todos al momento de analizar resultados, ya que puede cambiar una situación que se presenta a priori como un gran éxito, en un fracaso rotundo.

A continuación veremos algunos ejemplos para aclarar este concepto.

Caso I. El consumo de carbón y su relación con las enfermedades respiratorias

Este caso fue presentado a principios del siglo XX en

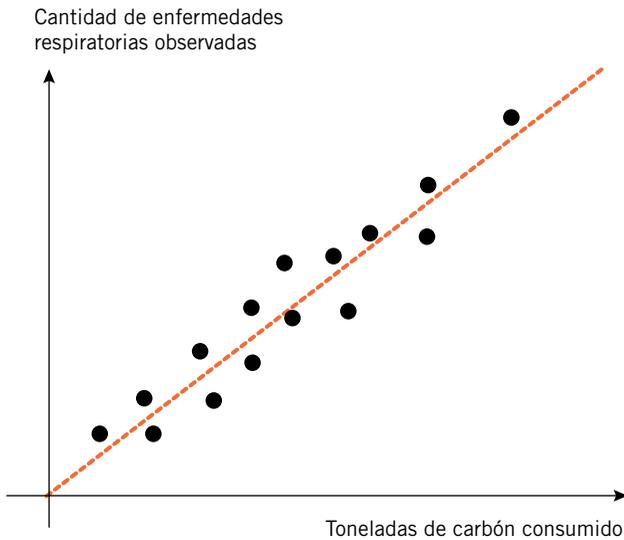


Figura 3.

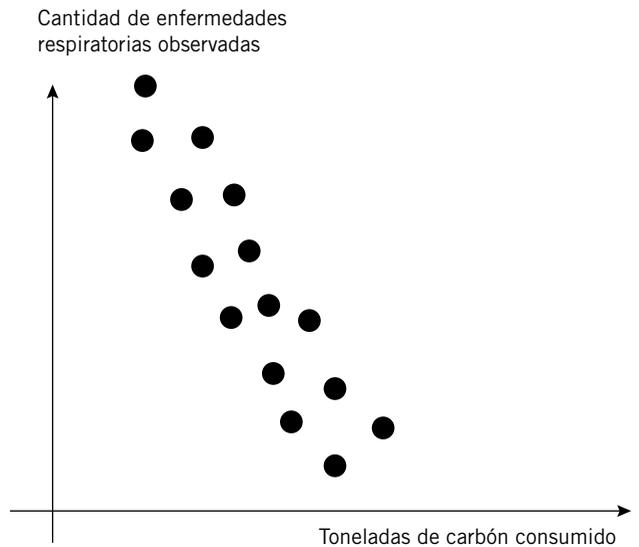


Figura 4.

Londres, en los albores de la estadística y de la visualización de información, los investigadores entusiasmados con las técnicas y los avances al momento pudieron observar una relación que podría ayudar a la definición de políticas públicas de salud. El resultado obtenido versaba más o menos así: “Hemos logrado determinar que el incremento en el consumo de carbón es un factor de alto impacto en la salud pública de nuestra región. Nuestros científicos han determinado que dicho incremento produce un incremento en la cantidad de enfermedades respiratorias observadas”.

Por supuesto que una conclusión de tal impacto fue acompañado de un gráfico que “prueba” dicha ley (Figura 3).

Fue una decisión política de aquel entonces, normar restrictivamente el consumo de carbón en la región. El resultado fue algo inesperado: hospitales y sanatorios sobrecargados de gente con enfermedades respiratorias (Figura 4).

El motivo de ese fracaso fue que los analistas no contemplaron las relaciones causales inherentes al problema y no consideraron un factor endógeno a la problemática tratada: la temperatura ambiente.

Un análisis contextual que atribuya correctamente las relaciones causales inherentes hubiesen dado un veredicto distinto: al bajar la temperatura ambiente se produce un incremento en el consumo de carbón debido a las necesidades de calefacción mientras que, al mantener cerrados los ambientes debido al frío, las defensas se reducen y la calidad del aire que se respira se empobrece aumentando el número de contagios de enfermedades respiratorias. En la figura 5 se muestra un diagrama causal que explica la relación entre las variables que generó la correlación espuria observada en los datos.

Caso II. El origen de un mito pintoresco

Siguiendo el mismo razonamiento, pero unos años antes, en Inglaterra, algunas observaciones indicaron que en las localidades en las que se observaban más cigüeñas, también se observaban más nacimientos.

La conclusión pintoresca fue el mito conocido por todos: a los niños los traen las cigüeñas.

Por supuesto que un análisis causal rápidamente pudo descartar a la relación espuria existente: En ciudades más grandes, hay una mayor cantidad de nacimientos (hay más personas). A su vez en ciudades más grandes hay más altillos y cúpulas, sitios en los que anidan las cigüeñas y, por ende, en ciudades más grandes habrá más cigüeñas.

Será entonces una tarea conjunta de los científicos de datos y de los analistas del negocio analizar los resultados para establecer las relaciones de causalidad y transformar así los patrones encontrados en conocimiento accionable. Esta es la finalidad de la implementación de técnicas de *analytics* en una organización.

Lo inmenso, lo rápido y lo complejo

Con el advenimiento de las nuevas tecnologías, en los últimos años se han producido y almacenado en las organizaciones más datos que en toda la vida de nuestra civilización. El ritmo de generación, producción y aceleración en la generación de nuevos datos es vertiginoso. Conti-

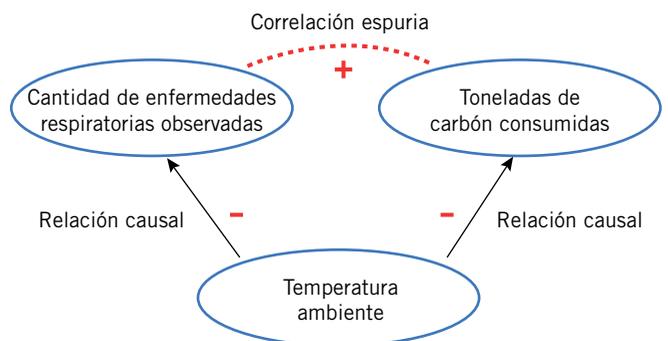


Figura 5.

nuamente salen nuevos sitios, aplicaciones y herramientas que producen datos.

Hasta hace un par de décadas, los datos históricos, por fuera de lo operativo, eran considerados “basura digital”. El surgimiento masivo del *analytics* cambió paulatinamente esta mirada y trajo aparejada la percepción de que esos datos no era basura, sino que podrían transformarse en un gran valor para la empresa si podía extraerse conocimiento de ellos. Entonces estos datos comenzaron a almacenarse, sus volúmenes crecieron exponencialmente y plantearon nuevos desafíos.

¿Cómo hacer entonces para no perder ese valor que posiblemente se encuentre oculto en ellos? ¿Cómo hacemos para explotar semejantes volúmenes de datos? ¿Cómo hacemos para procesarlos y para actuar con base a ellos cuando lo necesitamos o incluso en el mismo momento en que se producen?

El desarrollo cotidiano de nuevas tecnologías de información ofrece casi diariamente respuestas, mejoras y alternativas con el fin de mejorar estos puntos y resolver los problemas planteados. Con el objetivo de comunicarlos adecuadamente, las empresas y los desarrolladores les han puesto nombres pomposos: *Big Data*, *Real Time Analytics*, *Fast Data*, *Analytics* y CEP (Procesamiento de Eventos Complejos), entre otros.

Entendemos por *Big Data* al conjunto de tecnologías desarrolladas para trabajar con grandes volúmenes de datos, los que, debido a su heterogeneidad inherente, tienen grandes variabilidades que provienen de una gran variedad de fuentes de datos posibles, las cuales pueden ser estructuradas y no estructuradas, necesitando de una gran velocidad de procesamiento de manera tal de poder sustentar a soluciones operables dadas estas características.

Esta es una de las definiciones posibles para *Big Data* y se la denomina la definición de las cuatro V, por volumen-variabilidad-variedad-velocidad mencionadas en su definición. No es la única, pero es una de las más acuñadas y citadas en la bibliografía.

A modo de referencia, hablar de *Big Data* implica hablar de decenas de Gigabytes a Tera, Peta y/o hasta *Zettabytes* de datos a ser procesados por los sistemas.

Los términos *Fast Data Analytics*, *Real Time Analytics* y/o CEP apuntan a la temporalidad de los eventos y a su procesamiento **en el momento en que el dato es generado**. El valor de tomar una decisión en el momento en que el dato (oportunidad, problema, accidente, etc.) se produce suele ser mucho mayor que actuar a posteriori, ya que se minimiza el impacto de lo negativo y/o se gana por el efecto de una respuesta inmediata.

Un ejemplo claro es la ejecución de una acción correctiva ante la ocurrencia de un derrame de fluido y/o en una situación con riesgo de incendio antes y/o durante la aparición del fuego. Si a una herramienta de *Fast Data* (tal como TIBCO *Streambase*) le llegan las señales de determinados sensores, y las características medidas determinan una alta probabilidad de ocurrencia de un evento no deseado, la herramienta de *Fast Data* puede definir y/o ejecutar una acción que corte un flujo y/o detenga un bombeo, enviando una notificación de alarma a un operador y/o iniciando

alguna acción automática de sofocación si existen actuadores que permitan estas acciones. Otro caso de uso puede ser la detección temprana de una falla (en una electrobomba sumergible, por ejemplo), procesando la información proveniente de sensores, para alertar tempranamente a un operador y que el mismo inicie las acciones de mantenimiento predictivo y preventivo correspondientes. El valor aportado a una operación puede ser inconmensurable.

Metodología IVEMIS-ia®: Una estrategia orientada a la resolución de problemas basada en datos

A lo largo de mis 20 años como Científico de Datos me he encontrado con una gran cantidad de problemas a sortear en pos de implementar soluciones de manera eficiente y efectiva.

Uno de los problemas con los que me he encontrado en las implementaciones de *Analytics* es que el cliente no contaba con una cultura de análisis de información y producción de conocimiento. Al no contar con esta cultura, tampoco contaba con una metodología de trabajo a tal fin y menos aún con métricas que permitan al negocio medir los resultados obtenidos y cuantificar el grado de mejora obtenido.

En ese sentido, el primero de los puntos que debe comprender una organización es que *Analytics* es un **proceso** muy bien definido, con etapas claramente determinadas y que dicho proceso cumple con reglas generales que pueden sistematizarse en una metodología.

Metodologías marco hay varias. No todas son completas y algunas están sesgadas a solo una parte del proceso en sí. Presento aquí una metodología propia, desarrollada durante mis años de experiencia en la temática y que comparto públicamente bajo licencia GPL⁽¹⁾: “**IVEMIS-ia**®”.

IVEMIS-ia® es un acrónimo que resume los pasos de la metodología: “identificar-valorar-entender-modelar-im-

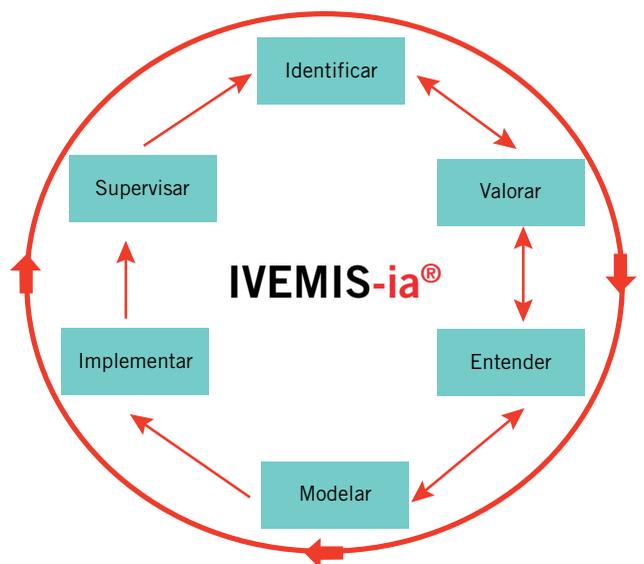


Figura 6.

plementar-supervisar”, cada una de estas etapas son partes constitutivas del proceso definido como un círculo de mejora continua (Figura 6).

En detalle, cada una de las etapas de IVEMIS-ia® mostradas en la figura 6 son las siguientes:

Identificar: detectar y/o tener la presunción de la ocurrencia de un problema en el negocio y/o en un proceso en particular. En esta etapa se debe detectar y/o presumir la ocurrencia de un problema por tratar y una vez hecho esto debe definirse al problema con claridad, mediante una declaración explícita.

Valorar: cuantificar el impacto de la resolución del problema. En esta etapa deben realizarse estimaciones sobre el impacto económico actual del problema (directo e indirecto), debe estimarse el valor aportado por una solución basada en *Analytics* a fin de poder fijar las expectativas en la solución plausible de ser obtenida y debe estimarse una relación costo/beneficio del proyecto en sí. En este punto se recomienda seguir un criterio de subordinación, tomando como primer objetivo la producción de valor y/o beneficio (por ejemplo, facturación), siguiéndolo por un pensamiento de reducción de costos y mirando por último al ROI del proyecto.

Entender: comprender y contextualizar el problema que será tratado. En esta etapa, los analistas deben determinar el GAP existente analizando la disponibilidad y la calidad de los datos, la disponibilidad de plataformas de almacenamiento, procesamiento y modelado analítico y el GAP “cultural y/o intelectual” de los RRHH de la organización. Deben evaluarse riesgos y mitigantes del proyecto, realizarse una adecuada contextualización del problema, comprenderse la naturaleza del mismo en cuanto a las reglas y/o naturaleza física del problema, las limitaciones existentes, las exclusiones necesarias de ser realizadas para la obtención de un modelo explicativo adecuado. También, deben determinarse la necesidad y la pertinencia (o no) de realizar muestreo sobre los datos, la estrategia de modelado y debe definirse la tarea analítica que será aplicada.

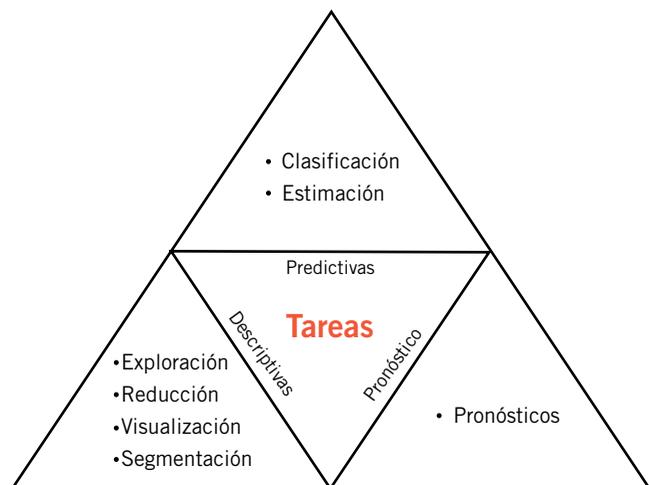


Figura 7.

En este punto resulta útil especificar que las tareas que ofrece *Analytics* pueden resumirse en tres grandes categorías como se muestra en la figura 7.

En esta etapa deben definirse el objetivo técnico que será analizado, preseleccionarse el o los algoritmos que se utilizarán, determinarse el volumen de datos necesario para el modelado, definirse las métricas de evaluación junto a un criterio que determine el éxito o no de un modelo en particular, además deben explorarse los datos disponibles y definirse de manera precisa la estrategia de procesamiento de datos que se llevará a cabo.

Como puede apreciarse, esta etapa es clave en relación con el éxito de un proyecto de *Analytics*.

Modelar: desarrollar y validar la solución analítica. En esta etapa debe implementarse la estrategia de procesamiento de datos integrando las fuentes de datos necesarias, construyendo atributos, categorizando variables, aplicando filtros y exclusiones adecuadas al fin perseguido.

Es aquí donde previo al modelado deben seleccionarse las muestras para entrenamiento y testeo de los modelos, ajustarse los algoritmos a los datos y, por último, validarse los resultados.

En este punto es preciso aclarar que las validaciones de un modelo deben tener en cuenta varios factores, a saber:

- Ajuste y precisión: capacidad de reducir la incertidumbre.
- Poder de discriminación y elevación: capacidad de separar.
- Ordenamiento: capacidad de un modelo probabilístico de ordenar los datos.
- Estabilidad: reproducibilidad entre distintas muestras. Poder de generalización.
- Interpretabilidad y/o coherencia semántica: confirmar una hipótesis causal.

El resultado de esta etapa es un modelo analítico que será utilizado para abordar la resolución del problema tratado.

Implementar: en esta etapa se pone en producción la herramienta analítica desarrollada, informándose los resultados de la misma.

Supervisar: en esta etapa se controla continuamente el correcto funcionamiento de la solución desarrollada, revisando la estabilidad, el poder de discriminación y de ordenamiento del modelo, así como su precisión. Se suelen tomar aquí decisiones relativas a cambios de puntos de corte en clasificación y/o a la necesidad de re-ajuste o re-desarrollo del modelo productivo. Esta etapa es la que se encarga de gestionar el ciclo de vida del modelo y/o del conocimiento adquirido, además, es el fin de la iteración del proceso y muchas veces el comienzo de una nueva iteración de valor.

Las tareas y técnicas del *Analytics*: una parte del abanico de posibilidades y de herramientas disponibles

Como he mencionado, en el momento de modelar es menester determinar primero que tarea analítica debe ser llevada adelante para la consecuente resolución del problema.

De manera general, los enfoques de *Analytics* pueden ser separados en tres grandes grupos: las tareas descriptivas, las tareas predictivas y las tareas de pronóstico.

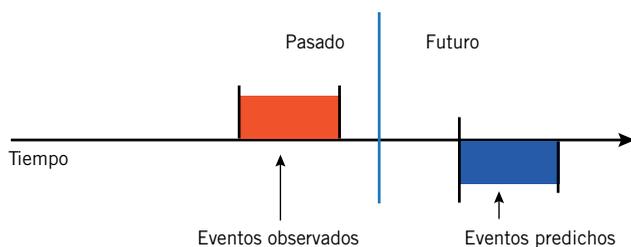


Figura 8.

Las tareas descriptivas o no supervisadas tienen como fin encontrar patrones y/o propiedades generales existentes en tabla de datos analizada y como fruto de estas tareas descriptivas se espera descubrir características importantes ocultas en los datos de un problema y/o de un proceso.

Dentro del abanico de algoritmos disponibles para cumplir las distintas tareas descriptivas nos encontramos con los estadísticos descriptivos, los distintos gráficos y tablas en la visualización de la información, los algoritmos de reducción de dimensionalidad, como el PCA, el análisis factorial o el ICA. Los algoritmos de segmentación, como K-Medias, *Clustering* jerárquico y/o Mapas Auto Organizados (Redes Neuronales de Kohonen) o los algoritmos APriori o FP-Growth para problemas de asociación o de secuencias.

Un ejemplo de tarea descriptiva es la interpretación automática de un conjunto de variables petrofísicas obtenidas de archivos LAS (LOG FILES) para la determinación de Facies (electrofacies, por ejemplo) utilizando técnicas de segmentación.

Las tareas predictivas o supervisadas apuntan a explicar un comportamiento en particular (problemático o no). En estas tareas se entrena un modelo utilizando los datos recolectados con anterioridad sobre la ocurrencia de un problema y/o evento en particular, para hacer luego predicciones futuras sobre la ocurrencia de dicho problema y/u objetivo (Figura 8). De manera general, la predicción nunca será 100% precisa, importando mucho en estos casos el rendimiento del modelo analítico cuando es aplicado a nuevos datos.

El abanico de algoritmos predictivos disponibles es muy grande, y se amplía cotidianamente. Podemos hablar de regresiones múltiples, *splines*, redes neuronales o árboles de regresión para problemas de estimación o regresiones logísticas, árboles de clasificación, algoritmos de inducción de reglas, máquinas de soporte vectorial (SVM), algoritmos bayesianos (Naive Bayes, por ejemplo) y redes neuronales de clasificación, entre muchos otros.

Algunos ejemplos de modelos predictivos pueden ser la selección de pares pozo-capas, en la cual realizar un punzado y posterior fractura hidráulica, considerando que la misma tenga alta posibilidad de ser productivamente eficiente y/o la predicción del surgimiento de eventos de fallas y/o desperfectos en electrobombas sumergibles, teniendo la información de producción y de los sensores previa al surgimiento del evento.

Por último, las tareas de pronóstico implican una mirada a futuro y en las cuales la variable tiempo desempeña

un papel explícito en el desarrollo y la formulación del modelo en sí.

Como modelos de pronósticos podemos hablar de modelos de suavizado exponencial, modelos de *Holt Winters*, modelos ARMA/ARIMA, ARCH, GARCH univariados o multivariados, o redes neuronales de estimación, las cuales suelen funcionar bien en problemas de pronósticos.

Ejemplos de modelos de pronósticos pueden ser el *forecasting* de producción de un pozo, batería y/o yacimiento o bien el pronóstico de precios del barril de crudo, del consumo de MW hora en una fecha en particular y/o el precio de la soja en el mercado de commodities a una fecha determinada.

Es menester destacar en este punto, que en la selección de algoritmos para el modelado, siempre debe seguirse una regla fundamental: la regla de la navaja de Ockham.

Regla de la navaja de Ockham: en igualdad de condiciones, la explicación más sencilla suele ser la más probable y, por ende, es la que debe ser seleccionada.

Los nuevos roles en la empresa

Uno de los grandes problemas con que se encuentra una organización al iniciar un proceso de implementación de *Analytics* es el de los recursos humanos necesarios.

Estas disciplinas han traído aparejado nuevos retos desde este punto de vista a las organizaciones, ya que las áreas de recursos humanos deben lidiar con perfiles hasta el momento por ellos desconocidos y fundamentalmente pasar de un pensamiento de especialización a uno de multidisciplinariedad.

En *Analytics* deben convivir hoy perfiles de los más variados, los que son muy difíciles de encontrar en un solo colaborador.

A modo de ejemplo, el colaborador ideal en *Analytics* sería alguien que tenga fuertes conocimientos en ciencias duras (Física, Matemática, Estadística, Inteligencia artificial, etc.), amplios conocimientos en bases de datos y tecnología, un gran conocimiento del negocio y sus disciplinas asociadas. En resumen, un perfil que podríamos llamar: la figurita difícil.

La alternativa a esta "figurita difícil" es la de conformar un equipo multidisciplinario que conjugue entre distintos RRHH, las distintas capacidades y que sea liderado por un RRHH que gestione la sinergia entre los distintos roles.

Un equipo así puede quedar conformado, al menos, por un analista funcional, encargado de establecer el vínculo con el negocio, por científicos de datos quienes se encargaran del desarrollo de las soluciones basadas en *Analytics* y por un interlocutor que gestione la relación con IT (con los administradores de sistemas, seguridad informática y de bases de datos).

Como es comprensible, surge entonces un nuevo rol en el nivel gerencial o directivo y es el rol de quien debe gestionar a un equipo de estas características: el CAO (Acronismo de *Chief Analytics Officer*) o Gerente de Soluciones Analíticas y/u el CDS (Acronismo de *Chief Data Scientist*).

El CAO es el gerente y/o director que deberá:

- Gestionar la interrelación con las áreas de negocio trabajando y operando internamente con el fin de alcanzar una nomenclatura única, corporativa y transversal a todas las áreas, sectores y colaboradores (un lenguaje único dentro de la organización para referirse a la información, a los problemas y a sus soluciones).
- Gestionar la capacitación de los recursos humanos de manera que los usuarios del negocio, los recursos de IT y los científicos de datos hablen “en un mismo lenguaje” sobre las necesidades y la oferta de soluciones posibles.
- Gestionar los presupuestos del área, encargarse de la difícil tarea de minimizar la rotación de los científicos de datos (cada vez más demandados por cierto).
- Elevar propuestas de mejoras para incorporarlas en la estrategia de la empresa.
- Gestionar la implementación, una metodología para Analytics como estrategia para el agregado de valor en el nivel organizacional.
- Gestionar la difusión interna de las nuevas capacidades adquiridas por la organización encargándose de “vender y/o propagandizar” internamente los logros y los resultados obtenidos por el equipo, con el fin de acompañar el proceso de aprendizaje y de acelerar internamente adopción de *Analytics* a la cultura organizacional. ■

IAE, en la UAI y en la UB en temáticas de Analytics y Big Data. Además, se ha desempeñado como investigador y desarrollador de tecnologías en el Instituto Nacional de Tecnología Industrial (INTI), como Becario de Investigación en la Comisión de Investigaciones Científicas de la Provincia de Buenos Aires (CIC) y en la Fundación Techint/UBA.

(1) GPL: Licencia de uso libre con obligación de mención pública del autor de la metodología.

Estanislao Martín Irigoyen es Socio Gerente en INTELIGENTIA S.R.L., empresa dedicada al desarrollo e implementación de Soluciones Analíticas y representante regional de la Plataforma Analítica TIBCO Spotfire. Se ha desempeñado como analista en Inteligencia de Negocios, Inteligencia Artificial y Data Mining en el equipo de Exploración de Datos Internacional Health Services Argentina, como Consultor y Docente en Estadística y Data Mining en SPSS Argentina y como consultor en SAS Institute Argentina.

En su tarea como consultor ha implementado soluciones basadas en Data Mining y Analytics en una gran variedad de industrias, como Petróleo/Gas y Energía, Manufactura, Agro Negocios, Retail, Farmacéutica, Bancario/Financiera, Salud, Telefónicas, Seguros y Medios, etc.

Académicamente ha sido docente auxiliar en carreras de grado en la Universidad de Buenos Aires en materias de Ciencias físicas, Ciencias económicas y en materias de Data Mining e Inteligencia artificial. Ha sido docente en diversos programas de extensión y postgrado en el