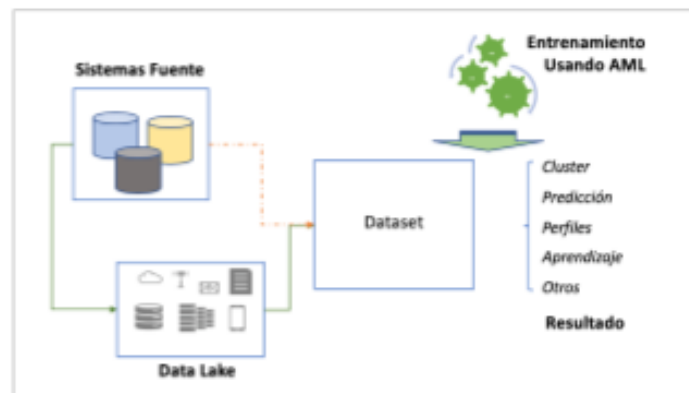


## DOS & TATIC DAN UN SALTO CUÁNTICO EN ANALÍTICA Y APRENDIZAJE DE MÁQUINA

Continuando con la discusión expuesta en el artículo 001, ahora se va a profundizar sobre:

- Visión y propuesta sobre las técnicas actuales usadas en Automatic Machine Learning (AML),
- Dificultad de alinear los resultados con la estrategia.
- Nuevos componentes analíticos propuestos por **TATIC**.
- Mejoras que se pueden agregar a los proyectos.

A continuación, en los numerales A, B y C, se presenta una descripción no exhaustiva de los procesos usados en Analítica Avanzada.



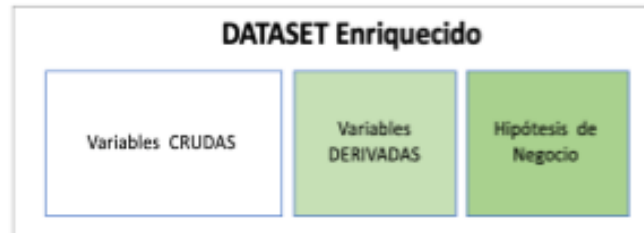
**A.** Conformar un “data set” con la mayor cantidad de columnas “crudas” provenientes de los sistemas fuentes, entre ellos: DWH1, ODS2, CRM3. En los últimos tiempos se incorporan a este proceso los Data Lake4.

**B.** Elegir y aplicar una técnica5 de Machine Learning (ML) adecuada para el problema que se quiere modelar, que se acople a los datos disponibles y saque el mayor provecho de estos.

**C.** Una vez seleccionada la técnica de ML, seleccionar el algoritmo o conjunto de algoritmos que mejor resultado de predicción o clasificación puedan brindar. Una buena técnica de hacer esta selección es usar la matriz de confusión.

Basado en la experiencia profesional, después de haber apoyado a muchas empresas en sus procesos de analítica avanzada, se propone hacer énfasis en el “armado” del data set y enriquecerlo con variables derivadas e hipótesis de negocio, definidas como: Situaciones de

negocio comunes a los problemas que se van a modelar y que ya han sido probadas en otros clientes e industrias.



Las variables derivadas son construidas desde variables crudas e incorporan conocimiento del negocio, incrementando el nivel de predicción del algoritmo de ML. En próximos artículos se presentarán ejemplos de casos reales donde el uso de variables derivadas evitó el obtener resultados obvios en los procesos de modelamiento predictivo – que se considera es la pesadilla de un científico de datos.

- Los resultados obtenidos a partir de las pruebas a las hipótesis de negocio permiten:
  - Desarrollar del enfoque de inteligencia artificial (IA) llamado búsqueda informada basada en Grafos.
  - Segmentar los clientes por situaciones de negocio comunes bajo una técnica desarrollada por **TATIC** denominada J. Boolean Ship, la cual se detallará en siguientes artículos.
  - Usar técnicas de aprendizaje reforzado, permitiendo que un mismo problema pueda ser abordado simultáneamente con el mismo data set bajo diferentes técnicas de ML supervisada y reforzada), aumentando con ello el desempeño del modelo.
  - En técnicas supervisadas, identificar las “fuerzas” Et (que mantienen los clientes o sujeto estudiado en condiciones de estabilidad) y las “fuerzas” Ep (que empujan a los clientes o sujeto fuera de su punto de equilibrio). Estos elementos serán detallados en subsecuentes artículos.

Estos nuevos componentes propuestos por **TATIC**, presentan mejoras comprobadas en los procesos de analítica avanzada y permiten que los resultados tengan alta adherencia con las acciones que las compañías deben desarrollar. • Derivación de variables con sentido de negocio. • Situaciones de negocio (hipótesis). • J. Boolean Ship. • Fuerza Et y fuerza Ep



**ALICE** es la novedosa propuesta analítica desarrollada por **TATIC**, usa Automatic Machine Learning, está pre construida y acelerada. Interactúa con soluciones de RPA (Robotic Process Automation) • **ALICE** Machine Learning Studio: Permite la creación de modelos de IA bajo los nuevos procesos analíticos desarrollados por **TATIC**.

- **ALICE** Customer Experience: Incorpora más de 200 variables derivadas y 150 hipótesis para modelar los siguientes casos: fuga de clientes, reclamo por servicio y/o producto, aceptación de una nueva oferta y solicitud de información. Todos estos modelos incluyen el cálculo del índice de margen de contribución.

**Definiciones:**

1) DWH: Un es repositorio unificado para todos los datos que recogen los diversos sistemas de información de una empresa y tiene propósitos analíticos y de reportes.

2) OSD: Es un repositorio unificado que generalmente solo almacena datos operacionales y tiene propósitos analíticos y de reportes.

3) CRM: Solución de seguimiento de las relaciones con los clientes, orientada normalmente a gestionar tres áreas básicas: la gestión comercial, el marketing y el servicio postventa o de atención al cliente.

4) Data Lake: Es un repositorio de almacenamiento que contienen una GRAN cantidad de datos estructurados y no estructurados en BRUTO, independientes de su fuente o formato. Actualmente, aunque no siempre, se crea sobre Hadoop.

5) Inteligencia Artificial (IA): Es un subsistema del campo de la informática que busca la creación de máquinas que puedan imitar los comportamientos inteligentes: conducir, analizar patrones, reconocer voces, jugar un juego y lograr un desempeño superior al humano.

6) Machine Learning (ML): Es una rama de la inteligencia artificial IA que busca dotar a las máquinas de capacidades de aprendizaje, es decir, generalización del conocimiento a partir de un conjunto de experiencias. Programar a una máquina para que pueda moverse es muy diferente a capacitarla para que aprenda a moverse. Algunas técnicas dentro del Machine Learning son: árboles de decisión, modelos de regresión, modelos de clasificación, técnicas para generar clúster y, tal vez la más nombrada, redes neuronales.



Artículo Analítica Avanzada 002-2019  
MAYO 8 DE 2019



Por: Jaime Barco Sierra - Director Regional de analítica – **TATIC**. -- Jaime.barco@tatic.net

[linkedin.com/in/jaime-barco-sierra-ba119380](https://www.linkedin.com/in/jaime-barco-sierra-ba119380)

---

**Belo Horizonte**, Av. Álvares Cabral, 1707 - 8º andar, Santo Agostinho | Belo Horizonte, MG | Brasil | 30170-915, Tel. +55 31 3024-0523, **Bogotá** Calle 26, No 69-76, Piso 16 Edificio Elemento. Torre 1, Bogotá | Colombia Tel +57 3155685182