CRISP-DM | Modeling: De los Datos Preparados al Conocimiento

Después de **entender el negocio**, **entender los datos** y **prepararlos**, llegamos a una de las etapas que más posibilidades abre dentro de un proyecto de datos: **Modeling**.

Quizás lo primero que debemos aclarar es que **modelar no significa necesariamente construir inmediatamente un modelo de Machine Learning o de Inteligencia Artificial**. Modelar también puede significar analizar.

Podemos comenzar con estadística descriptiva, distribuciones, relaciones entre variables, segmentaciones y patrones. Después podemos avanzar hacia análisis estadísticos más complejos, forecasting, modelos predictivos, Machine Learning o soluciones de IA.

La pregunta fundamental del negocio

Antes de elegir una técnica existe una pregunta mucho más importante: **¿Qué pregunta del negocio estamos tratando de responder con los datos?**

El objetivo no debería ser encontrar un algoritmo para nuestros datos, sino encontrar la mejor manera de utilizar los datos para responder una pregunta del negocio. Para hacerlo, consideramos la siguiente estrategia:

Estrategia: Ir de lo simple a lo complejo

- **Describir:** Comprender el estado inicial de los datos.
- **Comparar:** Identificar diferencias y variaciones clave.
- **Relacionar:** Buscar asociaciones y patrones relevantes.
- **Hipótesis:** Formular y contrastar supuestos analíticos.
- **Explicar:** Determinar causalidades e interacciones.
- **Avanzar:** Predecir, clasificar, segmentar, proyectar u optimizar cuando el problema realmente lo requiera.

Cada análisis nos permite comprender mejor nuestros datos dentro del contexto del negocio, y ese conocimiento fortalece el siguiente paso.

Experimentación, comparación e iteración

Lo complejo debería demostrar que aporta más valor que lo simple. Un modelo sofisticado no es automáticamente un mejor modelo. Podemos comenzar estableciendo una referencia sencilla y posteriormente probar diferentes alternativas. Si aumentamos la complejidad, deberíamos poder explicar qué estamos ganando con ella.

Modeling implica **experimentar y comparar**. Durante el proceso es fundamental responder preguntas clave:

- ¿Qué alternativas evaluamos y qué hipótesis estamos probando?
- ¿Bajo qué supuestos funciona determinada técnica?
- ¿Cómo comprobaremos nuestros resultados y qué parámetros modificamos?
- ¿Por qué seleccionamos una alternativa y descartamos otra?

Aquí aparece otro concepto fundamental: **La Iteración**. Modeling difícilmente es un camino perfectamente lineal: analizamos, encontramos algo, aprendemos, ajustamos y volvemos a probar. Incluso puede ocurrir que descubramos que necesitamos regresar a la preparación de datos. Regresar no significa que el proceso fracasó; **significa que aprendimos algo nuevo**.

El rol de la Inteligencia Artificial y el criterio humano

En esta etapa la IA puede convertirse en un multiplicador de nuestras capacidades. Las matemáticas fundamentales permanecen, pero las técnicas, herramientas, criterios y perspectivas disponibles evolucionan constantemente. Difícilmente una sola persona puede representar simultáneamente todas las perspectivas necesarias (negocio, estadística, ciencia de datos, tecnología, riesgo, finanzas, operaciones).

La IA ayuda a ampliar esas perspectivas: cuestionar hipótesis, explorar alternativas, comparar técnicas, interpretar resultados y evaluar nuevas formas de analizar el problema.

Tener más capacidad para analizar no significa delegar nuestro criterio. La IA puede proponer, calcular, comparar y encontrar patrones, pero somos nosotros quienes debemos comprender el contexto, validar los resultados y determinar qué tiene sentido para el negocio.

El principio de **Human in the Loop** adquiere todavía más importancia mientras mayor sea nuestra capacidad analítica.

Documentación y trazabilidad del proceso

Modeling no debería terminar únicamente con un gráfico, una regresión, un forecast o un modelo de Machine Learning. Debe dejar evidencia del proceso a través de:

- **Un documento de resultados:** Que explique qué hicimos, qué encontramos, cómo interpretamos los resultados y qué significan frente al objetivo planteado.
- **Una bitácora del proceso:** Que conserve la trazabilidad de nuestro razonamiento analítico (preguntas, hipótesis, técnicas, supuestos, prompts, alternativas evaluadas, descartes y validaciones finales del humano).

El conocimiento no está solamente en el resultado final; **también está en el camino que recorrimos para llegar a él**. Las organizaciones verdaderamente **Data Driven** construyen conocimiento organizacional a partir de cada iteración.

Valor de negocio sobre complejidad técnica

No deberíamos enamorarnos de un modelo ni confundir complejidad con calidad. Durante Modeling podemos determinar qué alternativa representa mejor el comportamiento de los datos, pero debemos recordar una regla de oro:

Un buen resultado técnico no significa automáticamente un buen resultado de negocio.

Podemos construir un modelo con excelentes indicadores técnicos o encontrar relaciones estadísticamente relevantes, pero la pregunta definitiva siempre será: **¿El conocimiento generado nos ayudó a tomar una mejor decisión y podemos demostrarlo?**

El verdadero propósito de Modeling es **transformar una pregunta de negocio en evidencia analítica mediante un proceso estratégico, progresivo, reproducible y documentado**. La combinación de metodología, IA y criterio humano es una herramienta extraordinariamente poderosa para lograrlo.

#CRISPDM • #Modeling • #DataDriven • #DataScience • #InteligenciaArtificial • #Analytics • #MachineLearning • #HumanInTheLoop