



Inteligencia de Negocios y Data Mining.

En capítulos anteriores ya revisamos lo que es Inteligencia de Negocios, y su relación con la Minería de Datos.

Actualmente los proyectos de Inteligencia de Negocios, son tanto un arte como una ciencia, requieren de conocimiento técnico y herramientas adecuadas, pero también requieren de experiencia y práctica.

Un proyecto de este tipo puede complicarse fácilmente, especialmente en lo que a Minería de Datos (o sus evolución: "La Ciencia de Datos") se refiere.

Los principales factores que afectan negativamente a un proyecto de este tipo son:

- Datos erróneos.
- Reticencia organizacional al cambio.
- Condiciones empresariales.
- Confundir causas con efectos.

Así mismo, el éxito de un proyecto de Minería de Datos depende de:

- Capacidad de medir el éxito de su implementación.
- Capacidad para monitorear el modelo.
- El costo de los errores.
- Experiencia.

A su vez, para poder llevar a cabo un proyecto de este tipo, se necesita de un grupo interdisciplinario con los siguientes conocimientos:

- Conocimiento del negocio.
- Conocimiento de los datos.
- Conocimiento de las técnicas de minería de datos.
- Conocimientos en la administración de proyectos.
- Capacidad para contar historias.

Por los motivos anteriores es conveniente seguir una metodología para el diseño e implementación de un proyecto de minería de datos, tomando en consideración que este tipo de procesos son continuos y en constante ajuste, para adecuarse a las necesidades de la empresa.



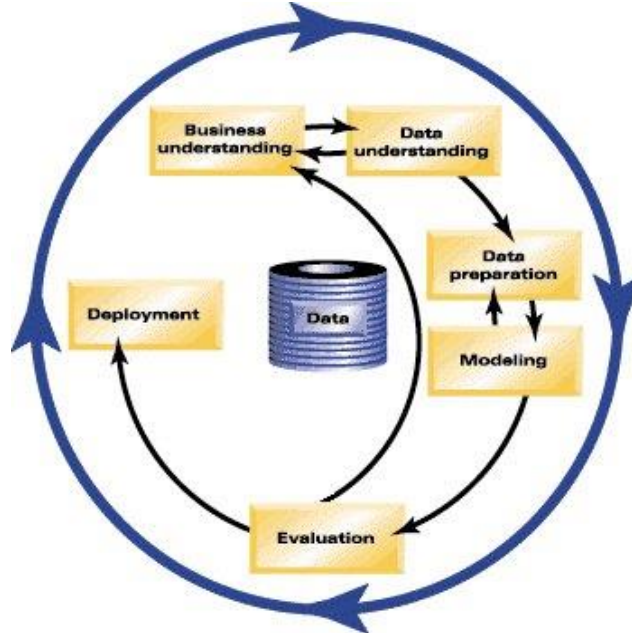
Metodología recomendada.

Si bien hay varias metodologías, la más aceptada por la industria es la Cross-Industry Standard Process for Data Mining, o CRISP-DM por sus siglas en inglés.

Este estándar considera 6 etapas en su desarrollo:

1. Comprensión del negocio.
2. Comprensión de los datos.
3. Preparación de los datos.
4. Modelado.
5. Evaluación.
6. Liberación.

En el siguiente diagrama se observa un poco más a detalle la relación entre estas diferentes etapas:



(Fuente: https://www.ibm.com/support/knowledgecenter/SS3RA7_17.1.0/modeler_crispdm_ddita/clementine/crisp_help/crisp_overview.html)

Cada una de estas etapas consta de tareas y subtareas:

1. Comprensión del negocio:

Tarea 1: Se determinan los objetivos del negocio: Es decir, se comprenden los antecedentes del ambiente, los objetivos de negocio, así como los criterios para determinar cuándo se considerará exitoso el proyecto.

Tarea 2: Analizar la situación actual: Hacer un inventario de los recursos, estimar riesgos y contingencias, así como ponerse de acuerdo en cuanto a la terminología.



Tarea 3: Determinar cuáles van a ser los objetivos de la minería de datos y, en su caso, su criterio de éxito.

Tarea 4: Generar un plan de proyecto, por escrito, identificando herramientas y técnicas para llevarlo a cabo.

2. Comprensión de los datos:

Tarea 1: Recolección y documentación las fuentes de los datos iniciales.

Tarea 2: Descripción documentada de los datos.

Tarea 3: Exploración de los datos y reporte de los mismos.

Tarea 4: Verificar la calidad de los datos.

3. Preparación de los datos:

Tarea 1: Seleccionar datos, usando criterios y razones de peso para incluir o excluir datos.

Tarea 2: Limpieza de datos.

Tarea 3: Generar los atributos derivados, así como agrupaciones de los mismos.

Tarea 4: Dar formato y combinar los diferentes conjuntos de datos, asegurándose de definir las unidades de análisis, e integrar adecuadamente los datos.

4: Modelado.

Tarea 1: Seleccionar las técnicas de modelado a usar, considerando las premisas y usos de cada modelo.

Tarea 2: Generar un diseño de pruebas. Es decir, estimar que datos, o porcentaje de los datos van a ser usados para "entrenar" el modelo y, dependiendo del modelo, algunos datos "testigo".

Tarea 3: Construir el modelo, definiendo los parámetros del mismo, así como la descripción del mismo.

Tarea 4: Probar el modelo, haciendo la revisión del mismo y, en su caso, ajustar los parámetros del mismo.

5. Evaluación.

Tarea 1: Evaluar los resultados, confrontando los resultados de los procesos de la minería de datos contra los criterios de éxito del negocio y, en caso de cumplir, aprobándolo.

Tarea 2: Revisión del proceso como un todo, es decir ¿se cumplió con los proyectos, tareas, responsabilidades y plazos para cada etapa?



Tarea 3: Determinar las siguientes acciones a seguir que, en caso de que los resultados o procesos no sean los que requiere el negocio, se deba regresar a una etapa anterior, establecer que si bien el modelo cumple con la descripción de los objetivos de negocio, pero no tiene la confiabilidad para ser liberado a producción, en cuyo caso se deberá revisar el modelo y ver si es posible mejorarlo o, en el mejor de los casos, determinar que se puede liberar a producción, pero continuando en un modelo de mejora continua, donde los modelos deban ser revisados periódicamente, para asegurarse de seguir satisfaciendo las necesidades de negocio para las cuales fueron creadas.

6. Liberación.

Tarea 1: Hacer un plan de liberación.

Tarea 2: Establecer un plan de mantenimiento.

Tarea 3: Realizar un reporte y presentación final del proyecto.

Tarea 4: Documentación del proyecto. Documentar la experiencia del proyecto, se recomienda hacer una presentación y un reporte de los resultados finales de todo el desarrollo, incluyendo tiempos, contratiempos y lecciones aprendidas.

Hay que recordar que el proceso de minería de datos es continuo, y que las etapas de la metodología no son lineales y pueden requerir de varias iteraciones.

Conceptos vistos:

Principales factores que afectan negativamente a un proyecto de DM.

Principales factores de éxito de un proyecto de DM.

Conocimientos necesarios para llevar a cabo un proyecto de DM.

Metodología CRISP-DM

- Comprensión del negocio.

- Compresión de los datos.

- Preparación de los datos.

- Modelado.

- Evaluación.

- Liberación.