

Universidad Castro Carazo

Portafolio de evidencias

Análisis de Datos

JUDITH BUSTOS SILES

Semana 2

Durante esta semana, en modalidad, asincrónica; realizamos la lectura de dos metodologías para la minería de datos, siendo una de ellas la principal y de donde derivan muchas otras más.

El primer concepto que logré definir fue minería de datos, y lo hice a través de la siguiente página web: https://www.sas.com/es_mx/insights/analytics/data-mining.html, donde en lenguaje sencillo se plasma la definición e importancia de dicho proceso. Continuando con las lecturas, de la primera de ellas, aprendí a identificar fuentes desde las cuales podemos obtener datos y al mismo tiempo, cómo esos datos requieren pasar por un proceso para identificar patrones. Dichos patrones pueden utilizarse bajo una metodología **CRISP-DM** (Cross-Industry Standard Process for Data Mining) y el cual consta de las siguientes etapas:

- **Comprensión del negocio**, el cual tiene como objetivo que a través de una serie de preguntas podamos conocer los requisitos de una solución desde la perspectiva empresarial.
- **Compresión de los datos**, durante esta etapa vamos a conocer el contexto de la solución, que se basa en técnicas estadísticas y de aprendizaje automático. Muchas veces se puede tener contacto con SME u otros departamentos, para conocer cómo fue la data recolectada, cómo es almacenada, etc.
- **Recopilación de datos**, etapa para identificar y reunir los recursos de datos disponibles, en caso de haber lagunas, se hace una revisión de los **requisitos de datos** y recopilación de más datos o nuevos datos.
- **Preparación de datos**, que abarca la limpieza de datos, combinar datos de múltiples fuentes y transformar los datos.
- **Modelado**, que se enfoca en desarrollar modelos predictivos o descriptivos según el enfoque analítico previamente definido.
- **Evaluación**, se evalúa el modelo para comprender su calidad y garantizar que abordar el problema empresarial de manera adecuada y completa.
- **Implementación**, que se llevará a cabo en un entorno de producción o en un entorno de pruebas comparables.
- **Retroalimentación**, se obtiene retroalimentación sobre el rendimiento del modelo y su impacto en el entorno en el que se implementó.

Junto a esta metodología de CRISP-DM, tenemos también **BADIR**. BADIR tiene como base ciertos procesos de la primera metodología mencionada. Sus siglas son la mención de sus cinco etapas:

1. **Preguntas del negocio** (Business Questions): importancia de conocer el contexto del negocio, identificar un problema a través de pregunta como ¿qué?, ¿cómo?, etc., ayudarán a encontrar más rápido un camino hacia la solución.
2. **Plan de análisis** (Analysis Plan): incluye cinco partes que son las metas, hipótesis, metodología, especificidad de los datos y un plan de proyecto.
3. **Colección de datos** (Data Collection): donde se realiza la recolección de datos y al mismo tiempo la limpieza de los mismos.
4. **Obtener información** (Derive Insights): es aplicar los tres pasos anteriores, tomar acción sobre ellos, ejecutar la metodología hacia la solución.
5. **Recomendaciones** (Recommendations): finalmente la etapa donde se recibe retroalimentación por parte de los clientes y la mismas son puestas en ejecución.