A screenshot of a computer screen

Description automatically generated

**Analítica para negocios**

Obtener conocimiento valioso para la toma de decisiones estratégicas (en el negocio) a partir de un análisis matemático-estadístico-algorítmico de datos.

A diagram of a data processing process

Description automatically generated

**Data understanding**

La fase de entendimiento de la información de CRISP-DM involucra tomar una vista de cerca a la información disponible para minería. Este paso tiene una gran importancia para la preparación de la información, que es la parte más larga del proceso.

El entendimiento de la información involucra acceder a la información y explorarla usando tablas y gráficos que pueden ser organizados usando la herramienta de proyección CRISP-DM.

**Pasos**

* **Recolectar datos**

La información puede provenir de diferentes fuentes, como:

**Información existente:** Esto incluye información como información transaccional, de encuestas, de entradas web, etc. Se debe considerar si la información existente es suficiente para satisfacer las necesidades.

**Información comprada:** Se puede considerar el uso de información suplementaria como la demográfica si esta es necesaria.

**Información adicional:** Si la información anterior no es suficiente, se pueden llevar a cabo encuestas o rastrear nueva información para complementar la ya existente.

**Lista de tareas**

¿Qué atributos (columnas) son las más prometedoras?

¿Qué atributos se pueden considerar irrelevantes y pueden ser excluidos?

¿Hay información suficiente para hacer conclusiones generalizables para hacer predicciones precisas?

¿Hay muchos atributos para su método de modelaje de su elección?

¿Estás uniendo varias fuentes de información? Si sí, ¿hay áreas que pueden presentar un problema al unirlas?

¿Has considerado cuántos datos faltantes se manejan en tus fuentes de datos?

Escritura de un reporte de recolección de datos: usando la información recolectada se

* **Descripción de los datos**

Hay muchas formas de describir datos, pero la mayoría de las descripciones se concentran en la cantidad y cualidad de los datos, cuánta información está disponible y la condición de los datos. Algunas características clave para describir datos son:

* **Cantidad de datos:** Para la mayoría de técnicas de modelado hay “intercambios” asociados al tamaño de los datos. A mayor cantidad de datos se pueden tener modelos más precisos, pero pueden tardar más el tiempo de procesamiento.
* **Tipo de datos:** los datos pueden tener diferentes formatos como numéricos, texto, booleanos. Prestar atención al tipo de dato puede sacar problemas en el modelado.
* **Esquemas de codificación:** frecuentemente los valores en las bases de datos son representaciones de las características como el género o tipo de producto, por ejemplo, una base de datos puede usar M para hombres y F para mujeres.

**Escribiendo un reporte de descripción de datos**

Para proceder efectivamente con el proceso de minería de datos considera el valor de producir una descripción precisa de los datos usando las siguientes métricas:

**Cantidad de datos**

¿Cuál es el formato de los datos?

Identificar el método usado para conseguir los datos.

Que tan grandes con las bases de datos, en número de columnas y filas.

**Calidad de datos**

¿Los datos incluyen características relevantes para el negocio en cuestión?

¿Qué tipos de datos se representan? (símbolos, números, etc.)

¿Calculaste estadísticas básicas para los atributos clave? ¿Qué conocimiento provee esto al negocio en cuestión?

¿Puedes priorizar atributos relevantes? Si no, ¿están disponibles analistas de negocio para proveer un mejor conocimiento?