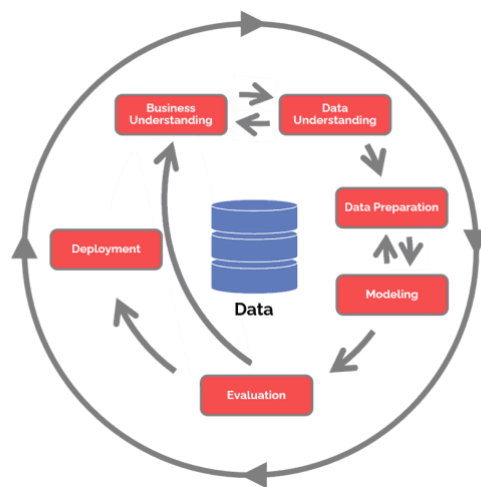


CRISP-DM

INTRODUCCIÓN.

El modelo CRISP-DM (Cross Industry Standard Process for Data Mining) indica los pasos que se deben seguir para construir un proyecto relacionado con la minería de datos. Este proceso, se ha venido puliendo a lo largo de los años, y este muestra 6 fases que todo investigador debe tener muy presente a la hora de realizar dicho proyecto. Los pasos a seguir son los siguientes:



1. Business Understanding phase (fase de entendimiento del negocio).
2. Data Understanding phase (fase de entendimiento de los datos).
3. Data Preparation phase (fase de preparación de los datos).
4. Modeling phase (fase de modelado).
5. Evaluation phase (fase de evaluación del modelo).
6. Deployment phase (fase de despliegue).

Cabe mencionar que este proceso es circular; esto quiere decir que el proyecto puede ir en proceso de evaluación y si hay que hacer algún reajuste, este puede volver, incluso, al proceso del entendimiento del negocio.

En este documento, se busca realizar la fase del entendimiento de un negocio de interés personal con el fin de simular un modelo CRISP-DM.

2. FASE DE COMPRESIÓN DEL NEGOCIO/PROBLEMA.

En este proyecto de minería de datos, se trabajará en un proyecto donde se trata de preveer el dinero pagado por el gobierno americano a usuarios por negligencia existente entre los años 1988 a 1997, utilizando el algoritmo de run-off triangle (que se trabajará en una fase posterior).

COMPRESIÓN DEL NEGOCIO

Se evidencia de acuerdo a estudios previos que uno de los índices más altos de dineros pagados a usuarios es gracias al sistema salud. En muchos casos, el usuario logra demostrar la existencia de procesos médicos que no son adecuados, acompañado de un mal servicio que muestra la negligencia del sistema de salud. Por esto, es importante tener claridad en la proyección del dinero que el gobierno debe tener (creación de reserva) para, en un principio, responder por dichas demandas generadas por negligencia.

OBJETIVOS DEL NEGOCIO

- **OBJETIVO GENERAL:** Determinar la reserva que debe tener el gobierno americano para responder a los usuarios del sistema de salud.
- **OBJETIVOS ESPECÍFICOS:**
 - a. Discutir con diversos profesionales sobre los problemas existentes en las reclamaciones por fraude en el sistema de salud.
 - b. Determinar factores principales de fraude y el dinero destinado a pagar cada una de estas reclamaciones, generando un total por año y el tiempo (medido en años) del pago de la reclamación.
 - c. Obtener registros históricos del dinero pagado por el gobierno a los usuarios del sistema de salud, entre los años de 1988 a 1997.
 - d. Organizar los datos utilizando un software específico.
 - e. Limpiar los datos de la base creada por medio de la recolección de datos, con el fin de elegir las variables de interés.
 - f. Confrontar a otros profesionales con la base de datos construida con el fin de verificar si la información escogida es la correcta para continuar con la implementación.

- g. Implementar el modelo run-off triangle con el fin de predecir los datos, usando la forma train test.
- h. Confrontar con profesionales sobre los resultados obtenidos, con el fin de crear la mejor reserva (reserva óptima).

EVALUACIÓN DE LA SITUACIÓN

- INVENTARIO DE RECURSOS

Los recursos que se necesitarían para realizar este proyecto son:

- i) Datos estructurados.
- ii) Personal profesional.
- iii) Distribución óptima de los recursos. Para nuestro negocio, se considera el siguiente organigrama:

METAS DEL ANÁLISIS DE DATOS

La meta deseada en el momento de realizar el análisis de datos es obtener el valor la reserva óptima que debe tener el gobierno para pagar a usuarios por fraude en el sistema hospitalario. Esto es muy importante, ya que si la reserva es muy pequeña, se corre el riesgo de no cubrir con los pagos adquiridos en ese año, llevando a una crisis financiera profunda, posiblemente la quiebra pero, si la reserva es muy alta, el dinero queda quieto, generando pérdidas al gobierno.