Primera lectura

Conceptos de mineria

Catalina Chavarría

2023

La minería de datos es un campo de estudio que se enfoca en descubrir patrones, relaciones y conocimientos útiles a partir de grandes conjuntos de datos.

Este proceso se puede definir mediante los seis pasos básicos siguientes:

1. Definir el problema:

Este primer proceso es el más importante, tenemos que comprender y especificar el objetivo que se pretende lograr mediante el análisis de datos. Al definir el problema se establecen las bases para el resto del proceso de minería de datos.

Se debe tener claro el problema que guiará todas las etapas posteriores de la minería de datos, incluida la selección de datos, la elección de algoritmos y la interpretación de resultados.

Diagrama

Descripción generada automáticamente

1. Preparar los datos:

Se refiere al momento inicial de preparar los datos para comprender su estructura, características y distribución. En este proceso se debe de asegurar que los datos estén limpios, estructurados y listos para ser utilizados en los modelos y algoritmos de minería de datos.

Diagrama

Descripción generada automáticamente

1. Extracción de datos:

Es el proceso de recopilar, limpiar y obtener datos.

“La garantía de calidad ayuda a detectar cualquier anomalía en los datos, como la interpolación de datos faltantes, manteniendo los datos en plena forma antes de que se sometan a una extracción.” Cita textual (¿Qué es el minado de Datos o Data Mininig? Técnicas y pasos a seguir, 2021)

Diagrama

Descripción generada automáticamente

1. Creación de modelos

Este proceso es definir y construir representaciones estadísticas de los patrones y relaciones que se encuentran en los datos. Estos modelos se utilizan para hacer predicciones, clasificar datos y comprender mejor la información contenida en conjuntos de datos.

Antes de construir un modelo, es importante preparar los datos. Esto se puede ya sea limpiando los datos y o eliminando valores atípicos

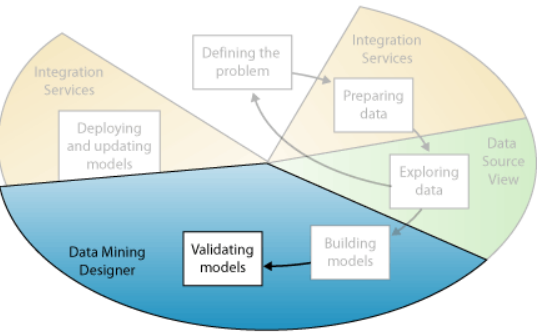
Diagrama

Descripción generada automáticamente

1. Exploración y validación de modelos

Es el proceso de identificar modelos inusuales o anómalas en los datos. Se utiliza para encontrar valores atípicos que puedan ser indicativos de errores o eventos inusuales.

“Antes de implementar un modelo en un entorno de producción, es aconsejable probar si funciona correctamente. Además, al generar un modelo, normalmente se crean varios con configuraciones diferentes y se prueban todos para ver cuál ofrece los resultados mejores para su problema y sus datos.” Cita textual (Conceptos de minería de datos, 2022)



1. Implementar y actualizar los modelos

Es importante evaluar y validar los resultados de la minería de datos para determinar su calidad y utilidad. Esto implica el uso de métricas de evaluación, técnicas de validación cruzada y pruebas estadísticas para medir la precisión, el rendimiento y la robustez de los modelos y algoritmos.

“Durante la implementación, otras partes interesadas utilizan el modelo de trabajo para generar inteligencia empresarial. El científico de datos planifica el proceso de implementación, que incluye instruir a otros sobre las funciones del modelo, realizar un seguimiento continuo y mantener la aplicación de minería de datos.” (¿Qué es la minería de datos?, s.f.)

Diagrama

Descripción generada automáticamente

# Bibliografía

*¿Qué es el minado de Datos o Data Mininig? Técnicas y pasos a seguir*. (2021). Obtenido de https://www.iebschool.com/blog/data-mining-mineria-datos-big-data/

*¿Qué es la minería de datos?* (s.f.). Obtenido de https://aws.amazon.com/es/what-is/data-mining/

*Conceptos de minería de datos*. (2022). Obtenido de https://learn.microsoft.com/es-es/analysis-services/data-mining/data-mining-concepts?redirectedfrom=MSDN&view=asallproducts-allversions&viewFallbackFrom=sql-server-ver16