

Fundamentos de análisis de datos: KDD (Knowledge Discovery in Databases)

Diferencias Principales:

En un inicio, consideraba que el proceso KDD y la Minería de Datos eran conceptos intercambiables, sin embargo, he llegado a comprender sus diferencias fundamentales. Desde mi perspectiva actual, el KDD representa una metodología integral que abarca todo el ciclo de trabajo con datos. Podría compararse con un proyecto completo que incluye múltiples etapas: inicia con la identificación y recolección de la información relevante, continúa con su preparación y limpieza para garantizar su calidad, y culmina con la interpretación y presentación de los resultados de forma comprensible y accionable. Por otro lado, la Minería de Datos y el Machine Learning constituyen componentes específicos dentro de este proceso más amplio. Estas técnicas representan la fase donde se aplican algoritmos especializados para descubrir patrones y extraer conocimiento de los datos ya procesados. Dicho de otra manera, mientras que KDD engloba la totalidad del proyecto desde su concepción hasta la entrega de resultados, la Minería de Datos corresponde a una etapa particular, aunque ciertamente crucial, donde ocurre el análisis y la extracción de información valiosa.

Fases Principales:

El proceso KDD se estructura en cinco fases secuenciales que serán analizadas en detalle: la Selección determina qué datos del repositorio original son pertinentes para los objetivos del análisis; el Preprocesamiento y Limpieza aborda la calidad de la información mediante la corrección de inconsistencias y el manejo de valores faltantes o atípicos; la Transformación adapta los datos a formatos compatibles con los algoritmos mediante técnicas como normalización o reducción dimensional; la Minería de Datos implementa los métodos computacionales que descubren patrones, asociaciones o modelos predictivos en la información preparada; y finalmente, la Evaluación e Interpretación valida estadísticamente los hallazgos y los contextualiza para convertirlos en conocimiento aplicable.

