

3. Operaciones básicas con BBDD

Las variables continuas son aquellas que pueden tomar un número infinito de valores dentro de un rango específico. Entre dos valores cualesquiera de una variable continua siempre existe otro valor posible. Por ejemplo, la temperatura, el peso y la altura son variables continuas porque pueden tomar cualquier valor dentro de un intervalo determinado. En el análisis de datos, las variables continuas permiten realizar operaciones matemáticas y estadísticas más complejas, como cálculos de medias, desviaciones estándar y correlaciones.

Por otro lado, las variables categóricas representan categorías o grupos discretos y no pueden dividirse en partes más pequeñas significativas. Estas variables son fundamentales para clasificar y organizar datos en grupos distintos que no tienen un orden inherente. Por ejemplo, variables como el género, el estado civil, o el tipo de producto son categóricas. Estas variables se utilizan comúnmente en el análisis de datos para identificar patrones y relaciones entre diferentes grupos. En resumen, las variables categóricas facilitan la agrupación y comparación de datos discretos, aportando una estructura clara y manejable al análisis.

La función del Feature Engineering se centra en la manipulación de características con dos finalidades principales: la interpretabilidad y la habilitación para su uso en Machine Learning. La interpretabilidad se refiere a la capacidad de los datos para ser comprendidos por los seres humanos. Al transformar y crear nuevas características, los ingenieros de datos buscan que las variables sean más intuitivas y fáciles de entender. Por ejemplo, convertir una fecha de nacimiento en una edad o calcular la diferencia de tiempo entre dos eventos son transformaciones simples que hacen que los datos sean más accesibles y comprensibles para los analistas. Durante la unidad didáctica desarrollaremos técnicas avanzadas de transformación que permitirán interpretar resultados a priori incomprensibles.

La habilitación para su uso en Machine Learning es la otra finalidad clave del Feature Engineering. Este proceso implica preparar y transformar los datos de tal manera que sean aptos para ser utilizados por algoritmos de aprendizaje automático. Esto incluye la normalización de valores, la codificación de variables categóricas o la creación de nuevas características que puedan capturar relaciones complejas en los datos. Por ejemplo, en un problema de predicción de ventas, generar características adicionales como el promedio de ventas por temporada o el impacto de promociones pasadas puede mejorar significativamente el rendimiento del modelo. La manipulación adecuada de las características permite a los algoritmos de Machine Learning aprender de manera más efectiva, mejorando la precisión y la robustez de los modelos predictivos.

En resumen, el Feature Engineering es esencial tanto para hacer los datos más comprensibles como para maximizar el rendimiento de los modelos de Machine Learning.

Las operaciones básicas en BBDD se refiere a las acciones fundamentales que se realizan para interactuar con los datos almacenados, con el objetivo de manipular y gestionar la información de manera efectiva.

Una base de datos es un conjunto de datos estructurados que pertenecen a un mismo contexto y se utiliza para administrar grandes cantidades de información de manera eficiente. Las bases de datos son esenciales para organizar, almacenar y recuperar datos, facilitando su gestión y análisis. Existen dos tipos principales de bases de datos: relacionales y no relacionales, cada una con sus características y ventajas específicas. A continuación resumimos ambos tipos:

- Las bases de datos relacionales organizan los datos en tablas estructuradas con filas y columnas, donde cada fila representa un registro único y cada columna representa un atributo del registro. Este modelo estructurado permite mantener la integridad y coherencia de los datos mediante el uso de claves primarias y relaciones entre tablas para establecer vínculos entre diferentes conjuntos de datos. Por ejemplo, una clave primaria en una tabla de "Clientes" puede estar vinculada a un atributo en una tabla de "Pedidos", estableciendo una relación entre un cliente y sus pedidos. Las bases de datos relacionales utilizan SQL (Structured Query Language) como lenguaje estándar para consultas y manipulación de datos, proporcionando una forma poderosa y flexible de interactuar con la base de datos.
- Por otro lado, las bases de datos no relacionales almacenan datos en estructuras más flexibles, como diccionarios o tablas jerárquicas. A diferencia del modelo tabular rígido de las bases de datos relacionales, las bases de datos no relacionales permiten un esquema más dinámico y adaptable, ideal para manejar datos no estructurados o semiestructurados. Este tipo de bases de datos es especialmente útil en aplicaciones que requieren escalabilidad y en entornos donde los datos pueden cambiar rápidamente, como las redes sociales o el Internet de las Cosas (IoT). Las bases de datos no relacionales ofrecen mayor flexibilidad al permitir almacenar diferentes tipos de datos en un solo lugar sin necesidad de definir estrictamente su estructura de antemano pero también son más difíciles de interpretar y explotar.

En resumen, tanto las bases de datos relacionales como las no relacionales tienen su lugar en la gestión de datos, cada una con ventajas específicas que las hacen adecuadas para diferentes tipos de aplicaciones y necesidades empresariales. En esta unidad didáctica nos centraremos en las bases de datos relacionales, explorando sus principios fundamentales, las operaciones básicas sobre ellas y la aplicación de la ingeniería de características.