What are common techniques for data transformation?

Data transformation is an important aspect of data analysis and processing. It serves as a link between raw data, and useful insights. Data transformation involves converting the data from its original structure or format into a more usable or suitable form. This process improves data quality and allows for compatibility with different systems. It also supports efficient analysis. Data transformation is a common step in data warehouses, integration projects, analytics pipelines and data warehousing. It is a crucial step to ensure that data is clean and consistent for downstream applications. **Data Science Course in Pune**

Normalization is a common data transformation technique that converts data into a standard range or format. This is particularly useful when data sets come from different sources, and have different scales or units. Normalization tasks include, for example, converting Fahrenheit temperatures to Celsius or standardizing date formats. Scaling numerical values to a fixed range (such as 0-1 or 0-1) can also improve the performance of certain machine-learning algorithms by eliminating magnitude bias. Normalization extends beyond categorizing data. For example, it can be used to convert free-text answers into structured labels.

Data aggregation is another widely used technique. It involves grouping or summarizing data to make analysis easier. In reporting systems, individual transaction records can be grouped into categories such as time, location or customer ID in order to generate meaningful metrics, like totals or averages. The aggregated data helps reduce the volume of data while maintaining essential characteristics for decision-making. As an example, the daily sales data could be aggregated to create monthly reports in order to identify trends over time.

Data transformation also includes filtering. This is the process of removing duplicates, irrelevant records, and erroneous data from a dataset. Analysts can improve the accuracy and reliability of the insights they derive from data by removing such data. Filtering is also possible based on conditional criteria, such as only selecting records that have missing values, or those from a specific geographical area or time period. It ensures that the analyzes are based upon relevant data of high quality.

Encoding is a crucial technique when dealing with categorical information. Categorical variables such as "gender" and "city" may be converted into numerical formats in order to be processed by algorithms requiring numeric input. This transformation is performed using techniques such as label encoding or one-hot encoding. Label encoding assigns a unique number to every category. One-hot encoding transforms each category into binary columns. This transformation allows algorithms for categorical data to be interpreted effectively during analysis and modeling.

Data integration is a technical transformation where data is combined from multiple sources into one format. It is crucial when working with heterogeneous platforms, like combining financial data from an SAP system with customer data from CRM. Integration may require mapping fields between different schemas and resolving discrepancies or conflicts in order to create a cohesive dataset. Data integration is essential for organizations to have a complete view of their business.

Data enrichment is the final transformative technique, whereby additional data are added to a dataset in order to increase its value. It could be adding demographic data to customer profile or integrating weather data in real time with sales records. The enrichment process can provide more context and uncover deeper insights not apparent from the original dataset. **Data Science Course in Pune**