Glossary

Module 1 Lesson 2: From Requirements to Collection

Welcome! This alphabetized glossary contains many of the terms you'll find within this lesson. These terms are important for you to recognize when working in the industry, when participating in user groups, and when participating in other certificate programs.

Term	Definition
Analytics team	A group of professionals, including data scientists and analysts, responsible for performing data analysis and modeling.
Data collection	The process of gathering data from various sources, including demographic, clinical, coverage, and pharmaceutical information.
Data integration	The merging of data from multiple sources to remove redundancy and prepare it for further analysis.
Data Preparation	The process of organizing and formatting data to meet the requirements of the modeling technique.
Data Requirements	The identification and definition of the necessary data elements, formats, and sources required for analysis.
Data Understanding	A stage where data scientists discuss various ways to manage data effectively, including automating certain processes in the database.
DBAs (Database Administrators)	The professionals who are responsible for managing and extracting data from databases.
Decision tree classification	A modeling technique that uses a tree-like structure to classify data based on specific conditions and variables.
Demographic information	Information about patient characteristics, such as age, gender, and location.
Descriptive statistics	Techniques used to analyze and summarize data, providing initial insights and identifying gaps in data.
Intermediate results	Partial results obtained from predictive modeling can influence decisions on acquiring additional data.
Patient cohort	A group of patients with specific criteria selected for analysis in a study or model.
Predictive modeling	The building of models to predict future outcomes based on historical data.
Training set	A subset of data used to train or fit a machine learning model; consists of input data and corresponding known or labeled output values.
Unavailable data	Data elements are not currently accessible or integrated into the data sources.
Univariate	Modeling analysis focused on a single variable or feature at a time, considering its characteristics and relationship to other variables independently.
Unstructured data	Data that does not have a predefined structure or format, typically text images, audio, or video, requires special techniques to extract meaning or insights.
Visualization	The process of representing data visually to gain insights into its content and quality.

Author(s)

Dr. Pooja
Patsy Kravitz

Changelog

Date	Version	Changed by	Change Description
2023-08-03	0.1	Patsy Kravitz	Initial version created