

词汇表：从理解到准备

欢迎！这个按字母顺序排列的词汇表包含了您在本课程中会遇到的许多术语。这些术语在您参与行业工作、用户组以及其他证书项目时非常重要。

术语	定义
自动化	使用工具和技术来简化数据收集和准备过程。
数据收集	从各种来源收集和组装数据的阶段。
数据汇编	组织和结构化数据以创建综合数据集的过程。
数据格式化	标准化数据以确保一致性和便于分析的过程。
数据操作	将数据转换为可用格式的过程。
数据准备	清理、转换和格式化数据以进行进一步分析的阶段，包括特征工程和文本分析。
数据准备	转换和组织数据以促进有效分析和建模的阶段。
数据质量	评估数据的完整性和完整性，解决缺失、无效或误导性值的问题。
数据质量评估	评估数据的完整性、准确性和完整性。
数据集	用于分析和建模的数据集合。
数据理解	数据科学方法论中的阶段，专注于探索和分析收集的数据，以确保数据能够代表待解决的问题。
描述性统计	数据科学家用来描述和理解变量分布的汇总统计，如均值、中位数、最小值、最大值和标准差。
特征	数据中的一个特征或属性，有助于解决问题。
特征工程	基于领域知识创建新特征或变量的过程，以提高机器学习算法的性能。
特征提取	从数据集中识别和选择相关特征或属性。
交互过程	基于数据分析的洞察和反馈对方法论进行迭代和持续改进。
缺失值	数据集中缺失或未知的值，在数据准备过程中需要谨慎处理。
模型校准	调整模型参数以提高准确性并与初始设计保持一致。
成对相关性	确定不同变量之间关系和相关性的分析。
文本分析	分析和操作文本数据的步骤，提取有意义的信息和模式。
文本分析分组	从文本数据中创建有意义的分组和类别以进行分析。
可视化技术	数据科学家用来创建视觉表示或图形的方法和工具，以增强数据模式、关系和洞察的可访问性和理解。

作者

[Dr. Pooja](#)
[Patsy Kravitz](#)

