



Data Preparation & Visualization

Data Pre-processing Techniques

Course: 23MCT305 - Data Analytics in Automation System

Faculty: N. KARTHI, AP/MCT



Improves Data Quality

Ensures accuracy and reliability



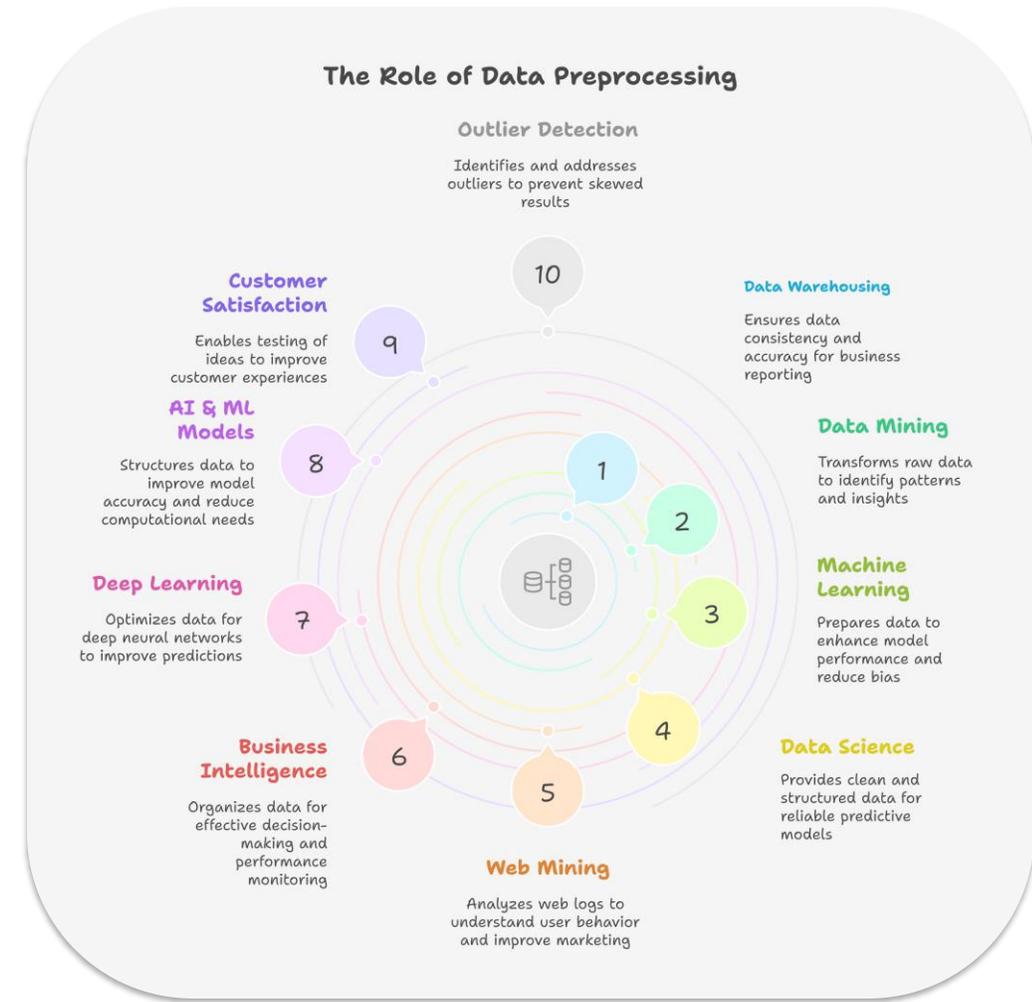
Enhances Model Performance

Boosts prediction accuracy



Ensures Data Consistency

Standardizes formats and values



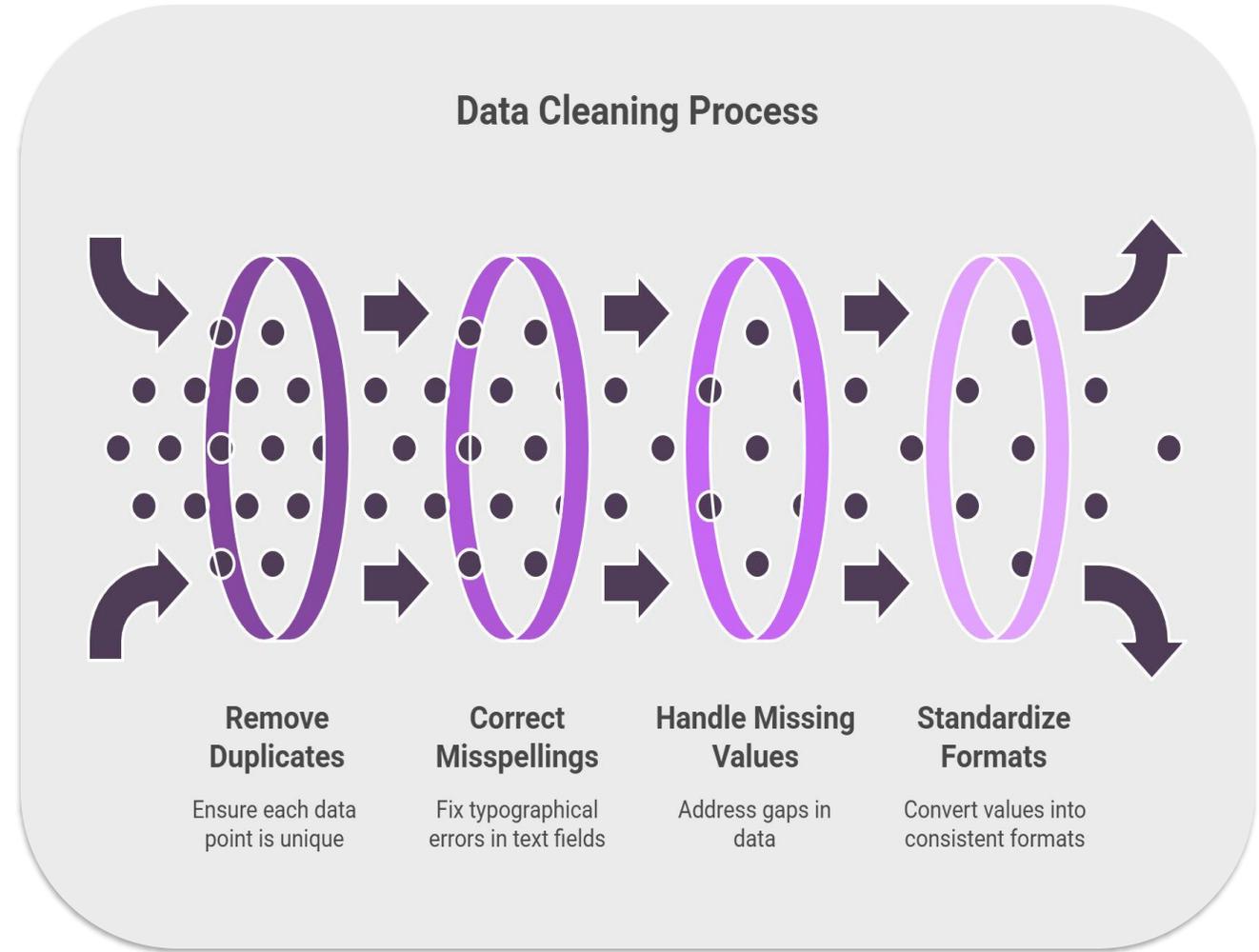
Key Stages

- 1 Data Collection
- 2 Data Cleaning
- 3 Data Transformation
- 4 Data Reduction
- 5 Data Integration
- 6 Data Visualization

Image failed to load:
<https://sfile.chatglm.cn/images-ppt/4777bf2617f.pn>
..

Key Techniques

-  **Handling Missing Values**
Imputation or removal
-  **Removing Outliers**
Statistical detection
-  **Eliminating Duplicates**
Deduplication methods
-  **Noise Reduction**
Smoothing techniques



Core Techniques



Feature Scaling

MinMax, Standardization



Data Encoding

Label, One-Hot Encoding



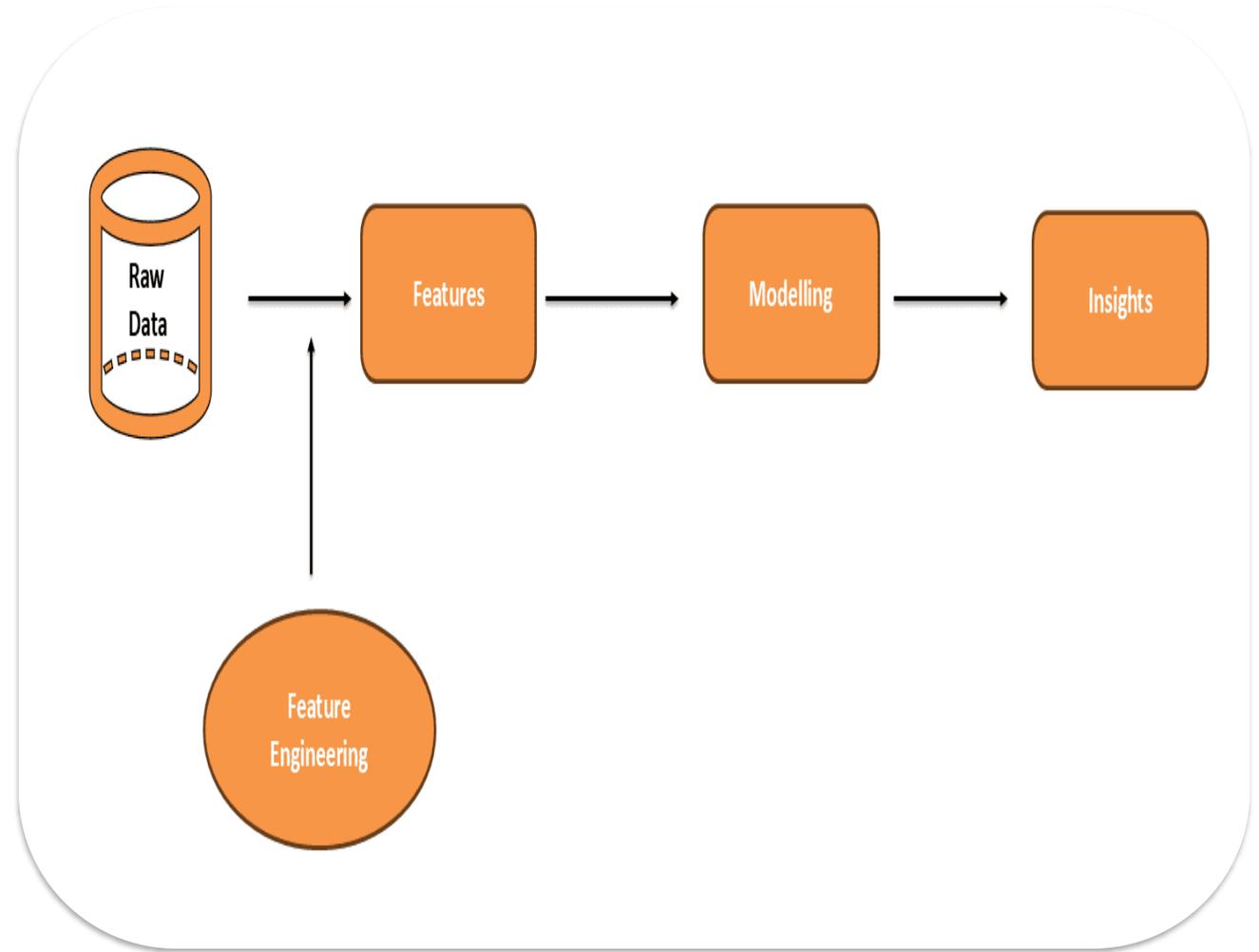
Feature Creation

Derive new features



Transformation Methods

Log, Square root, Power



Key Methods



Principal Component Analysis

PCA, t-SNE, UMAP



Feature Selection

Filter, Wrapper, Embedded



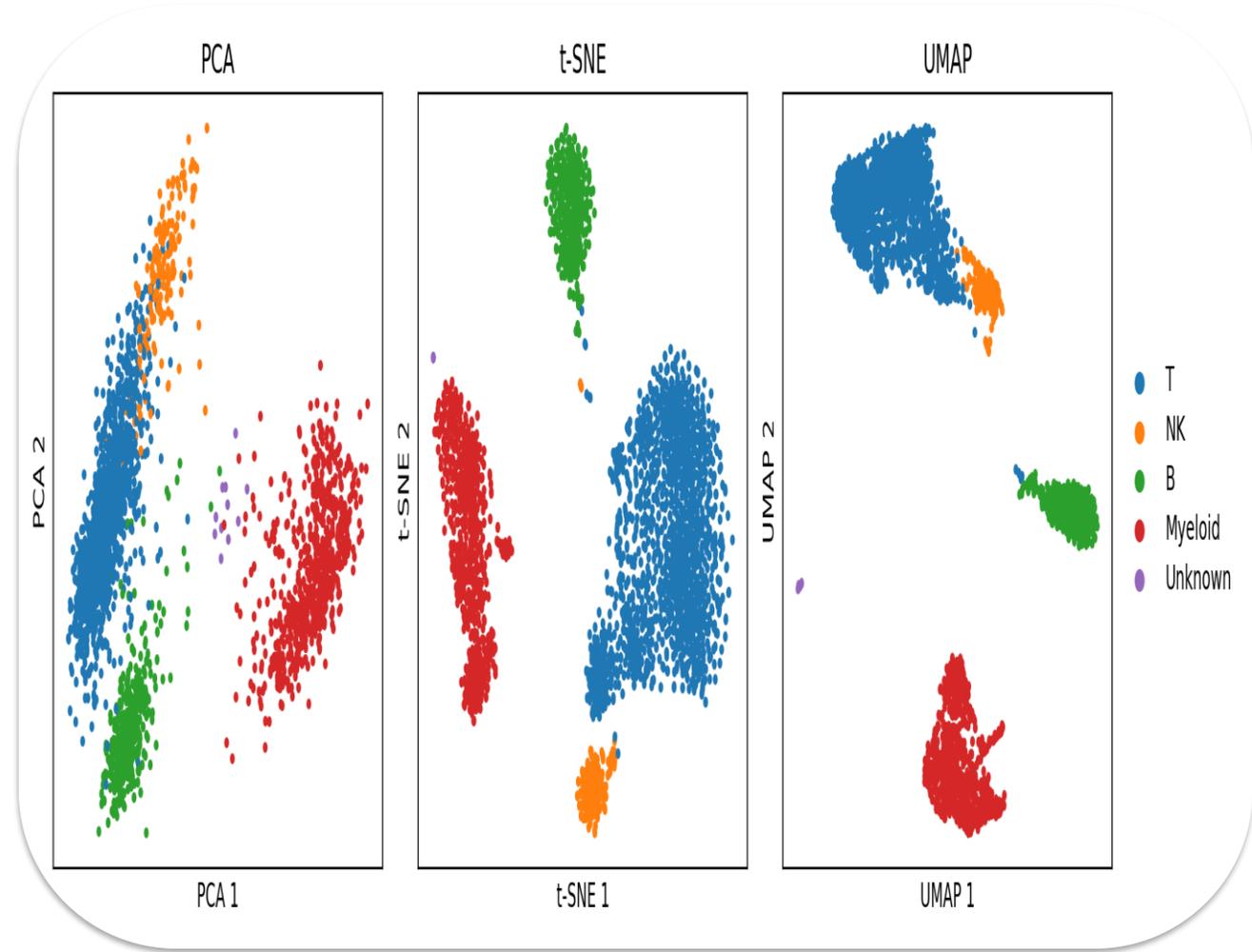
Sampling Techniques

Random, Stratified, Cluster



Dimensionality Reduction

Preserve variance, reduce complexity



Integration Strategies



Merging Datasets

Join, Union, Concatenation



Handling Conflicts

Schema matching, Resolution



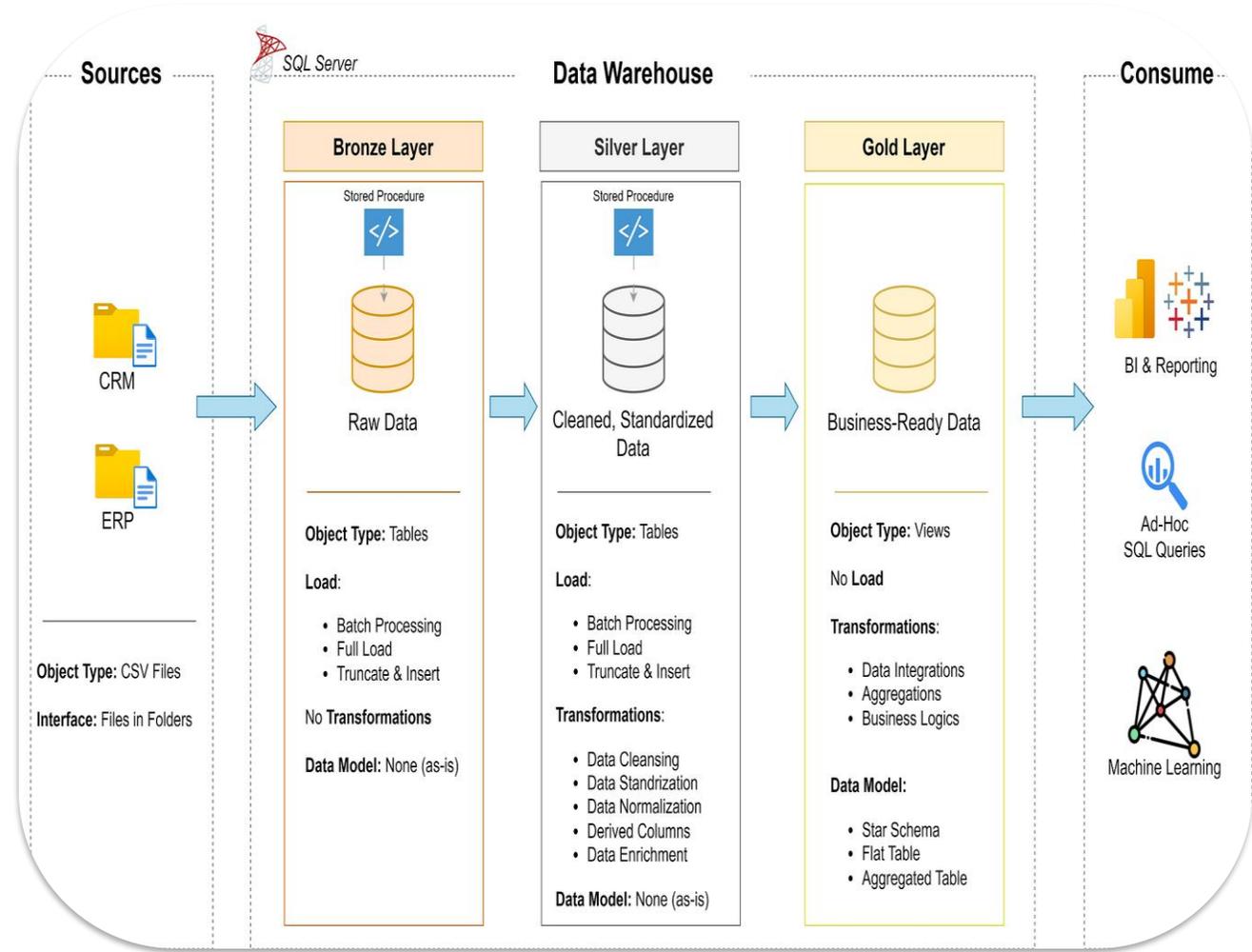
Data Warehousing

Centralized storage architecture



Multi-source Integration

ETL, ELT pipelines



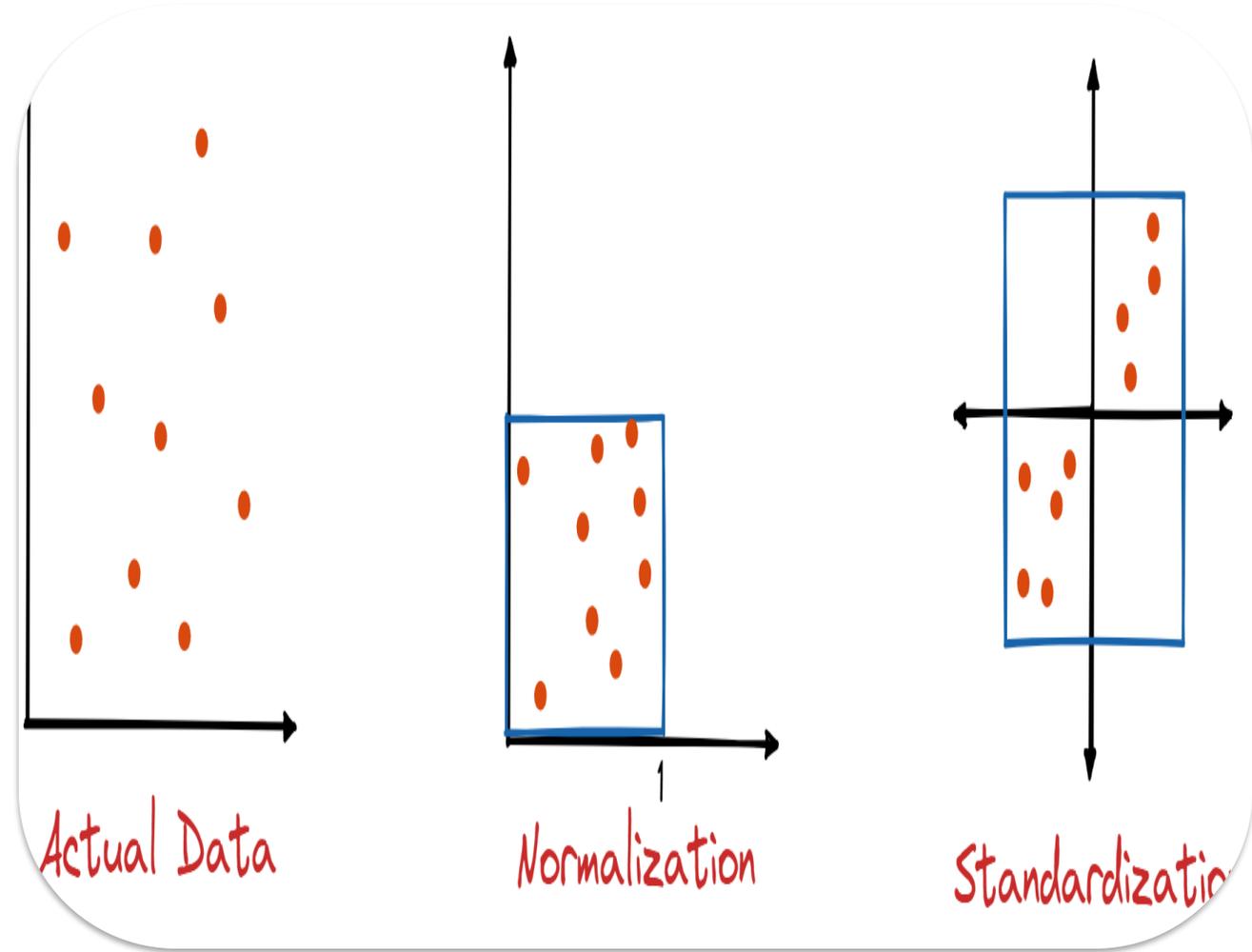
Scaling Methods

 **MinMax Scaling**
Range [0, 1] or [-1, 1]

 **Z-score Standardization**
Mean=0, Std=1

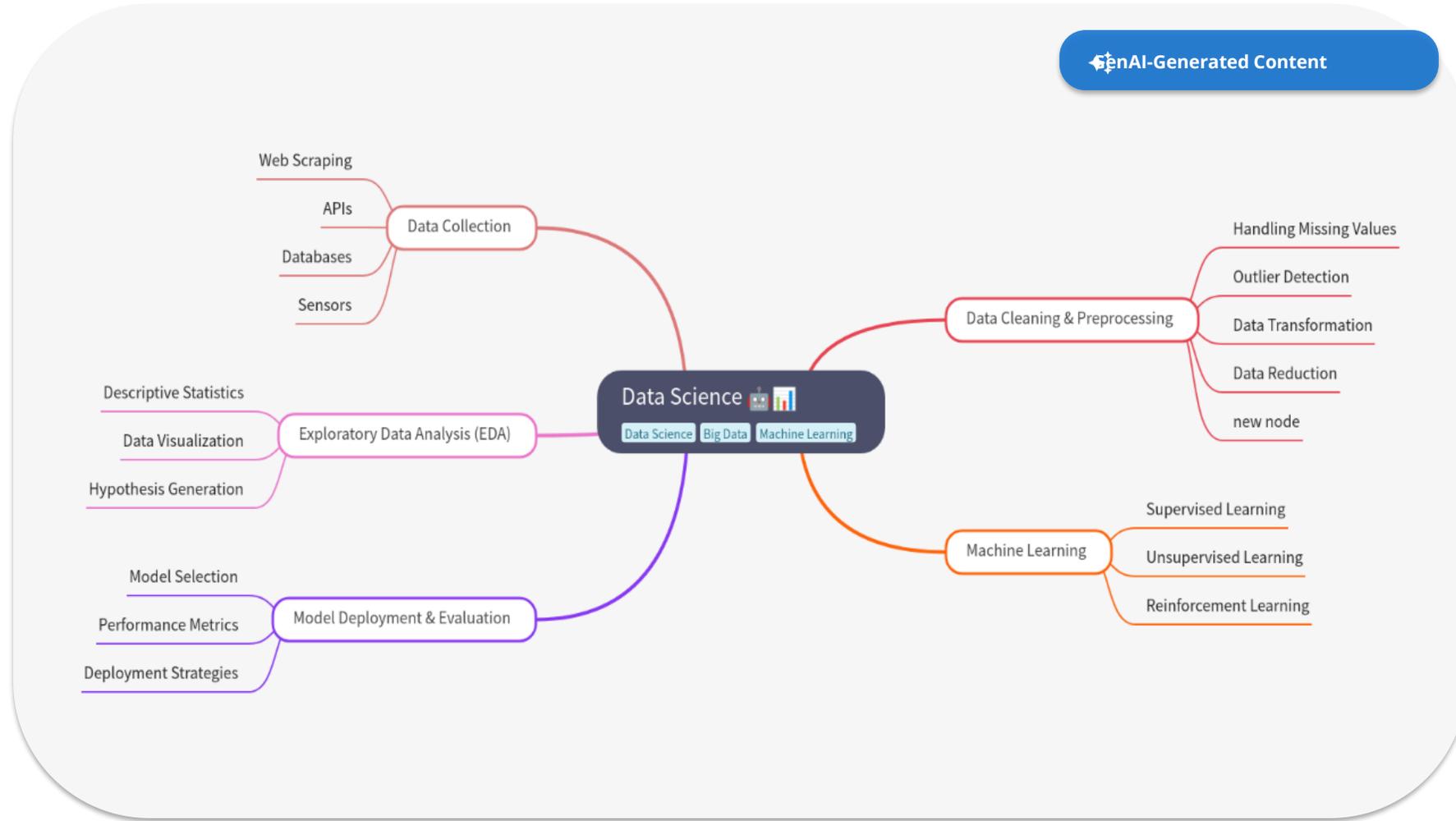
 **When to Use Each**
Dataset-dependent choice

 **Comparison Benefits**
Algorithm convergence speed





Mind Maps - Data Pre-processing Techniques



Key Takeaways

- **Data Quality Foundation**
Essential for accurate analytics
- **Systematic Workflow**
Structured pipeline approach
- **Multi-stage Process**
Cleaning, transformation, reduction
- **Continuous Improvement**
Iterative optimization cycle

Data and Analytics 3-Step Journey

