



Data Preparation & Visualization-Feature Extraction

Course: 23MCT305 - Data Analytics in Automation System

Faculty: N. KARTHI, AP/MCT

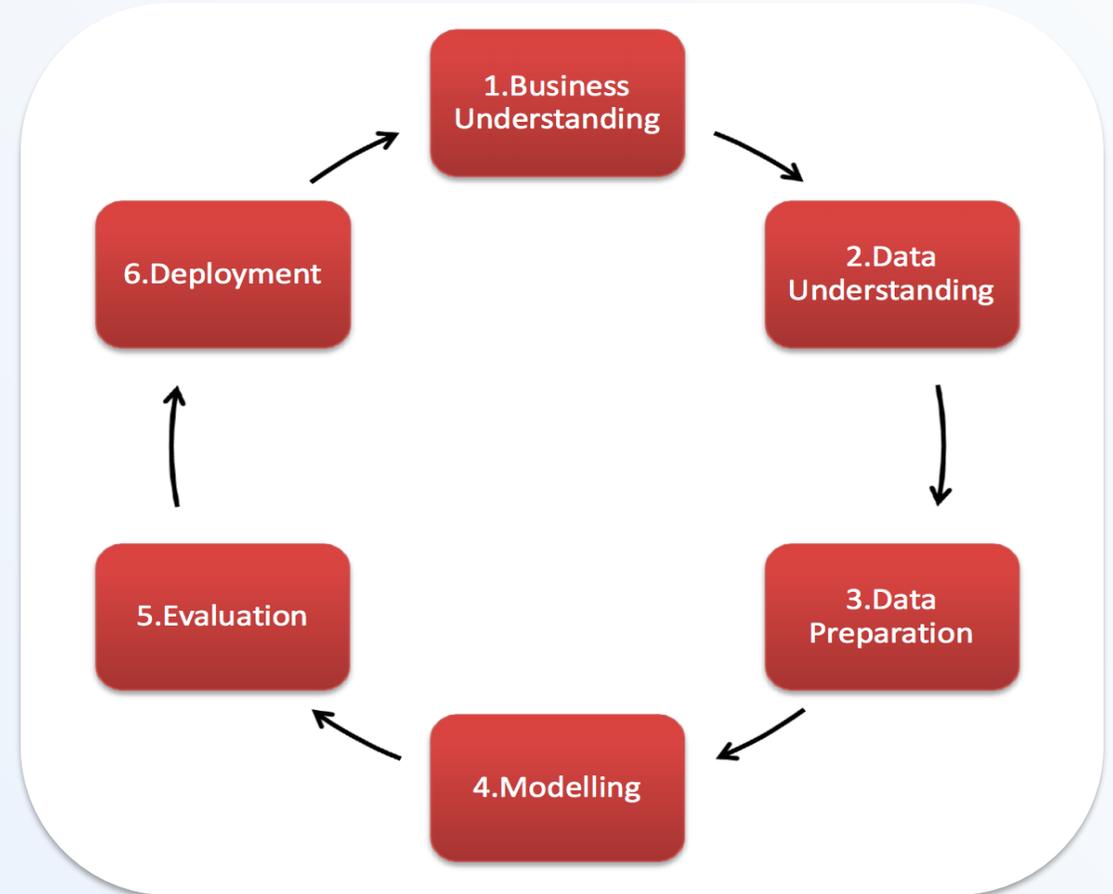
Introduction to Data Preparation



The process of cleaning, transforming, and organizing raw data for analysis and visualization.



- Ensures data quality and accuracy
- Reduces analysis errors
- Improves model performance
- Enables effective decision-making



Steps in Data Preparation



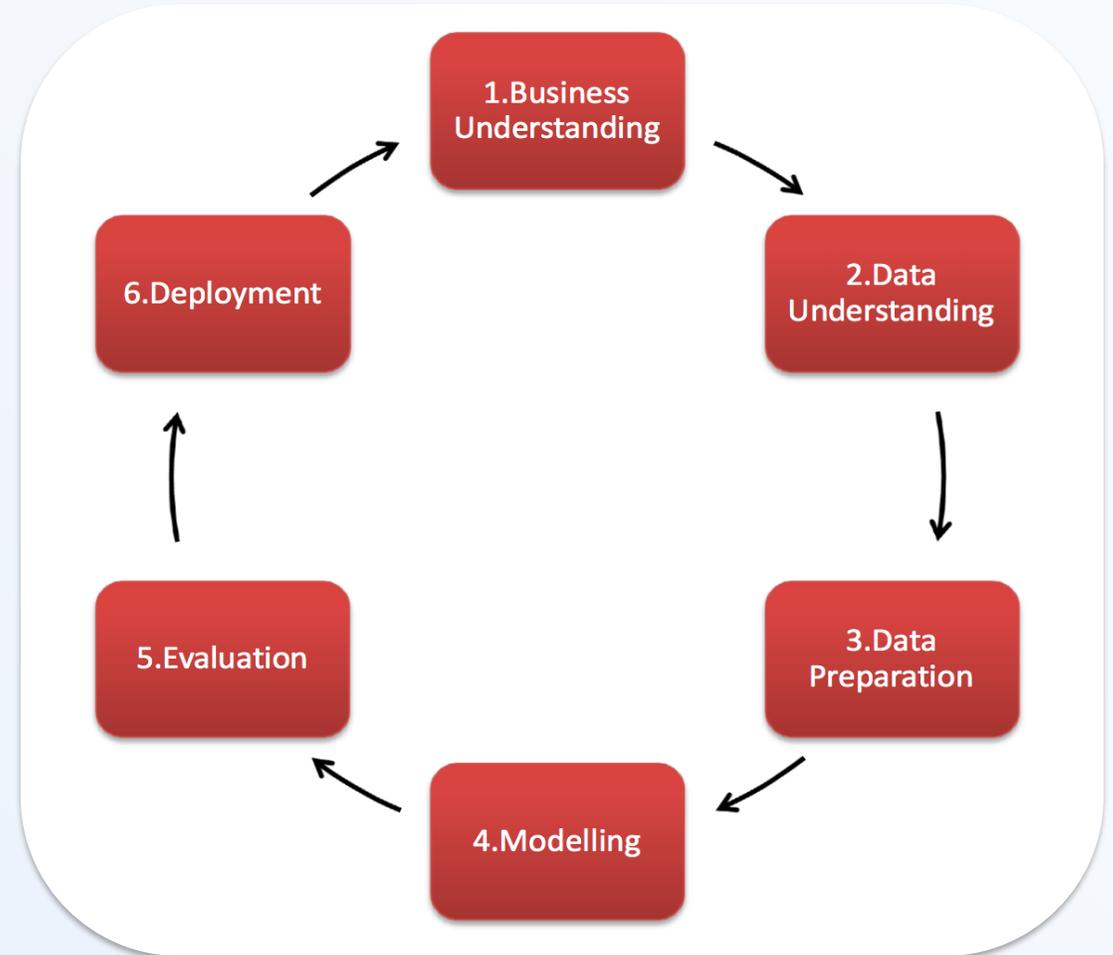
Gather raw data from multiple sources



Handle missing values and outliers



Normalize and scale features



Data Cleaning Techniques



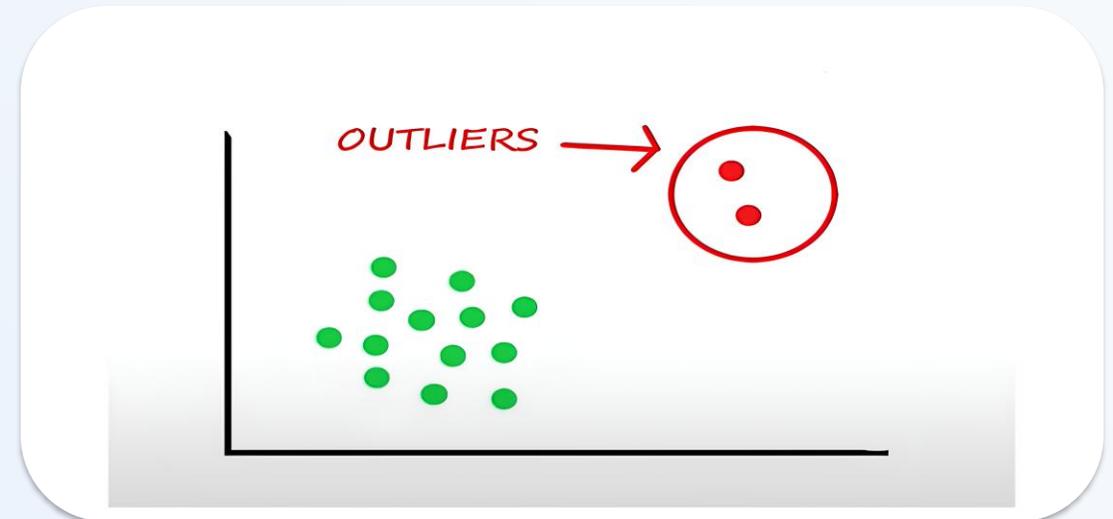
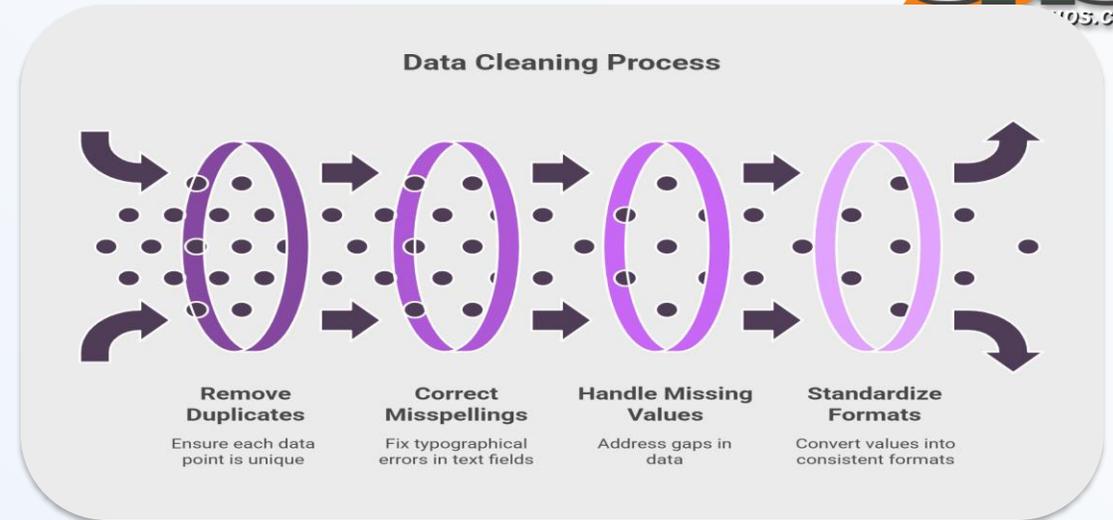
- Delete or impute missing entries
- Mean, median, mode replacement



- Identify unusual data points
- Use Z-score, IQR method



- Scale features to similar range
- Min-Max, Z-score scaling



Data Transformation Methods



- Normalize feature ranges
- Standardization techniques



- Convert text to numbers
- One-hot, Label encoding



- Reduce feature space
- PCA, LDA, t-SNE methods

DATA TRANSFORMATION TECHNIQUES



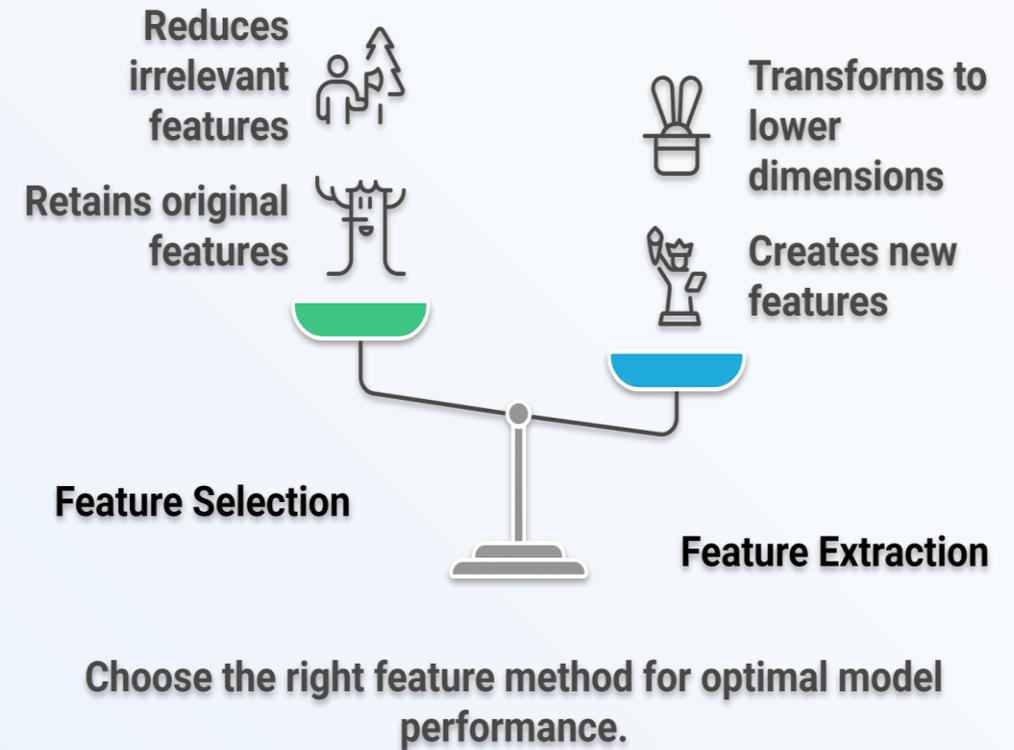
Introduction to Feature Extraction



Process of creating new features from existing data to improve model performance.



- Reduces dimensionality and complexity
- Enhances model accuracy
- Speeds up training process
- Reveals hidden patterns



Feature Extraction Techniques



Principal Component Analysis

- Linear dimensionality reduction
- Maximizes variance



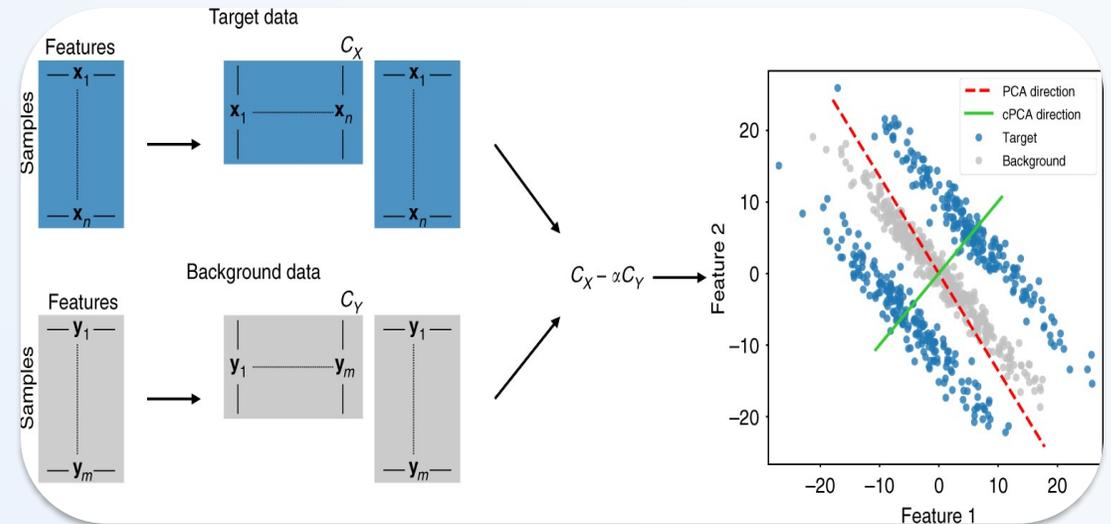
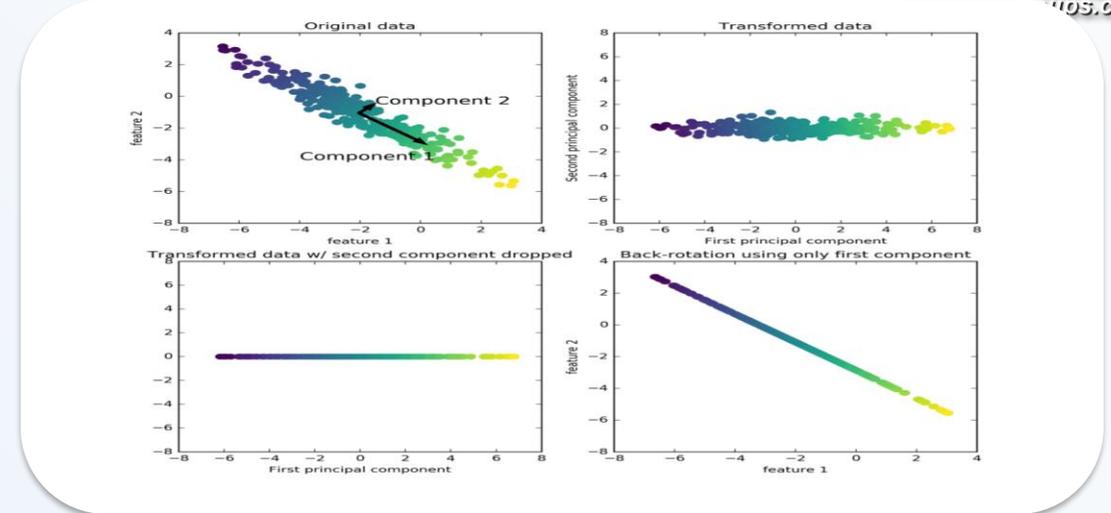
Linear Discriminant Analysis

- Supervised learning method
- Maximizes class separability



t-Distributed Stochastic Neighbor Embedding

- Non-linear technique
- Preserves local structure



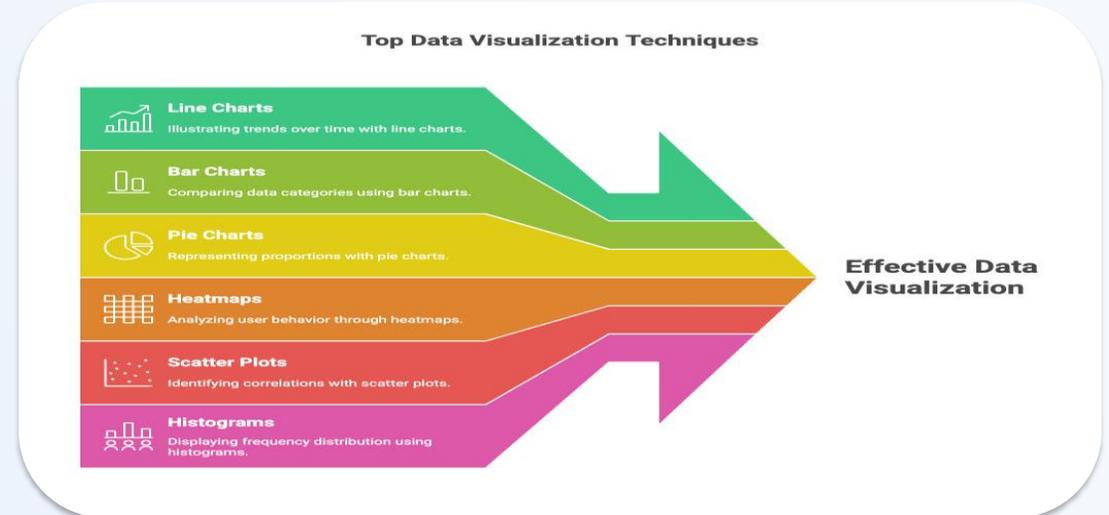
Data Visualization



- Makes data easier to understand
- Reveals patterns and trends
- Supports decision-making process

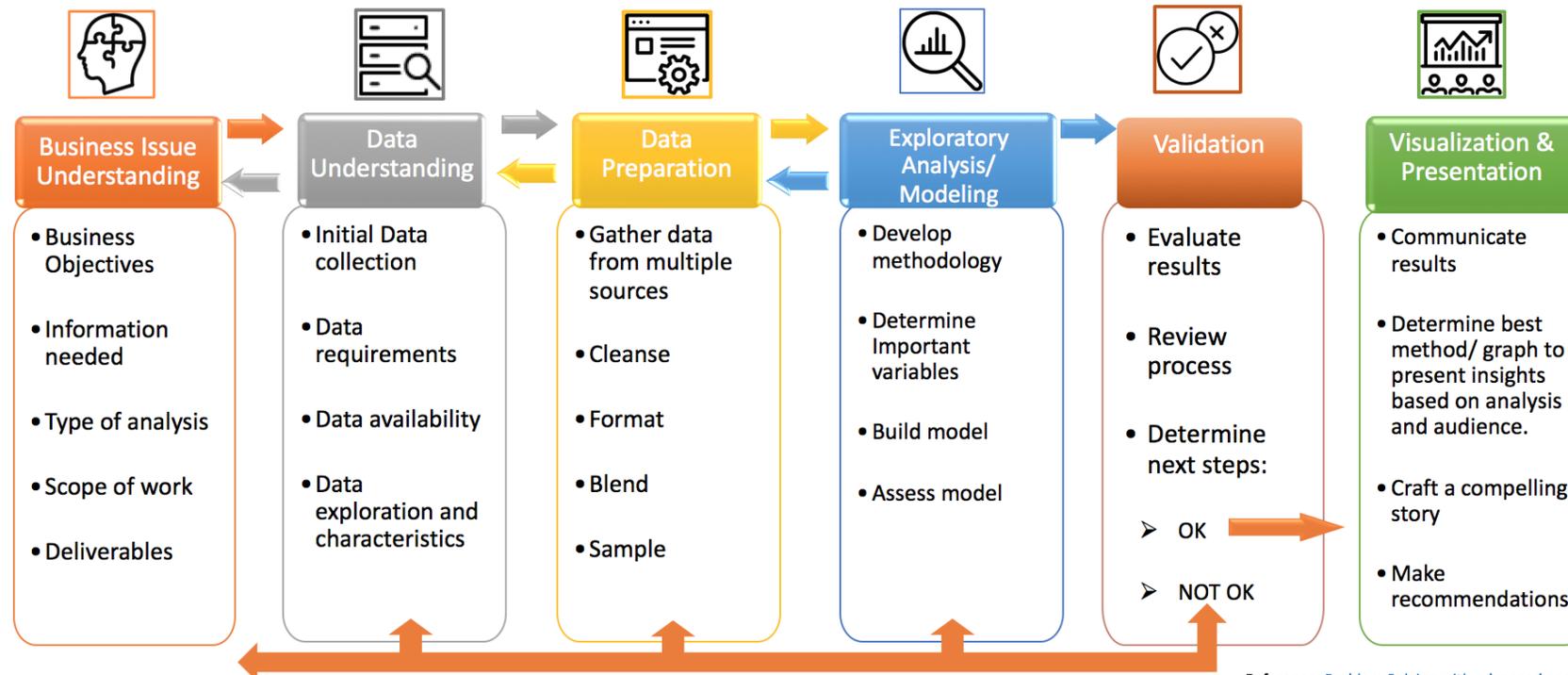


- Charts: Bar, Line, Pie
- Graphs: Scatter, Network
- Plots: Box, Heatmaps



Mind Map - GenAI Generated

Life Cycle of a data analysis project based on CRISP- DM methodology



Reference: https://en.wikipedia.org/wiki/Cross-industry_standard_process_for_data_mining

Recap and Summary



- Cleaning raw data
- Transforming data formats
- Handling missing values



- PCA: Maximizes variance
- LDA: Class separability
- t-SNE: Local structure



- Charts: Bar, Line, Pie
- Graphs: Scatter, Network
- Reveals patterns & trends

Complete Data Analysis Process

