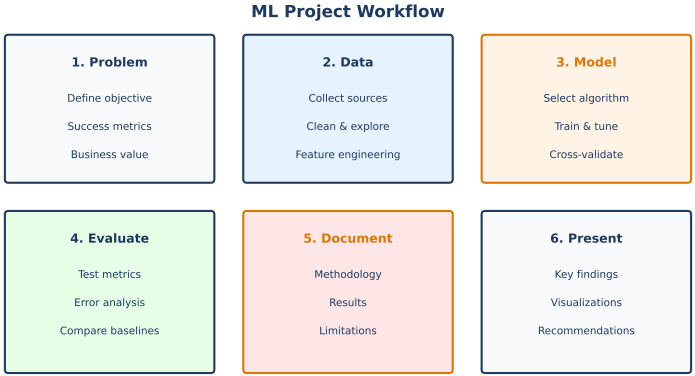
Lesson 45 Summary: Project Work 1

Data Science with Python – Key Concepts

Data Science Program

Project Workflow Overview



Structured approach from problem to presentation

Step 1: Problem Definition

Start with clarity:

- **Question:** What are you trying to predict/classify?
- **Metrics:** How will you measure success?
- **Value:** Why does this matter?

Clear problem = clear solution

Step 2: Data Preparation

Data is everything:

- **Collect:** Gather relevant data sources
- **Clean:** Handle missing values, outliers
- **Engineer:** Create meaningful features

80% of ML work is data preparation

Step 3: Model Development

Iterate and improve:

- **Baseline:** Start simple (logistic regression)
- **Experiment:** Try multiple algorithms
- **Tune:** Optimize hyperparameters

Simple models first, complexity when needed

Step 4: Evaluation

Honest assessment:

- **Metrics:** Accuracy, precision, recall, F1
- **Validation:** Cross-validation results
- **Comparison:** Beat the baseline?

Test set is the final judge

Quick Reference

Project Checklist:

Phase	Deliverable
Problem	Clear objective statement
Data	EDA notebook with visualizations
Model	Trained model with metrics
Evaluate	Test set performance report
Document	README + code comments
Present	Slides + demo

Follow the workflow for project success