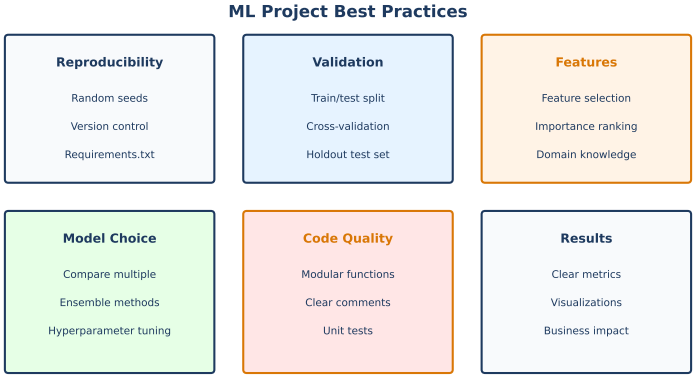
Lesson 46 Summary: Project Work 2

Data Science with Python – Key Concepts

Data Science Program

Project Best Practices



Professional ML project standards

Reproducibility

Essential for trust:

- **Seeds:** Set random seeds everywhere
- **Version:** Git for code, DVC for data
- **Environment:** requirements.txt/conda.yml

Anyone should get the same results

Code Quality

Professional standards:

- **Modular:** Functions, not scripts
- **Documented:** Docstrings, comments
- **Tested:** Unit tests for key functions

Good code = maintainable code

Model Selection

Systematic approach:

- **Compare:** Try multiple algorithms
- **Validate:** Use cross-validation
- **Tune:** Grid/random search

The best model depends on the problem

Documentation

Complete the story:

- **README:** Project overview, setup
- **Notebooks:** Exploratory analysis
- **Report:** Methodology, results

Document as you go, not at the end

Quick Reference

Project Essentials:

Component	Purpose
requirements.txt	Reproducible environment
README.md	Project documentation
src/	Production code
notebooks/	Exploration/analysis
tests/	Unit tests
models/	Saved model artifacts

Standard project structure enables collaboration