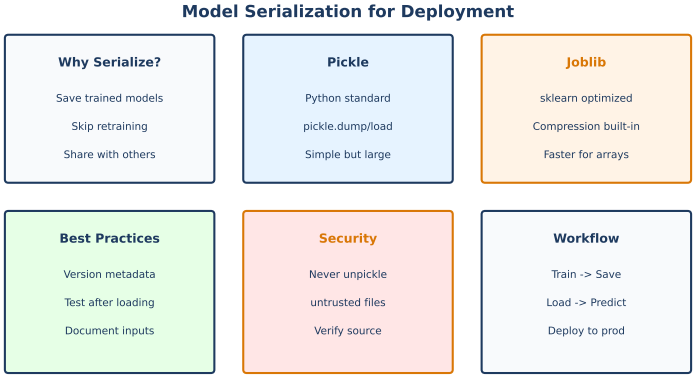
Lesson 41 Summary: Model Serialization

Data Science with Python – Key Concepts

Data Science Program

Model Serialization Overview



Save trained models for deployment

Why Serialize Models?

Key benefits:

- **Persistence:** Save trained models to disk
- **Efficiency:** Skip expensive retraining
- **Sharing:** Distribute to production/colleagues

Essential for any ML deployment pipeline

Pickle

Python standard library:

- **Save:** `pickle.dump(model, file)`
- **Load:** `model = pickle.load(file)`
- **Note:** Use binary mode ('wb'/'rb')

Simple but can produce large files

Joblib

Optimized for sklearn:

- **Save:** `joblib.dump(model, 'model.pkl')`
- **Load:** `model = joblib.load('model.pkl')`
- **Compression:** `compress=3`

Recommended for sklearn models with large arrays

Security Warning

Critical security risk:

- **Never** unpickle untrusted files
- **Pickle** can execute arbitrary code
- **Verify** model source before loading

Only load models from trusted sources

Quick Reference

Essential Commands:

Task	Code
Pickle save	`pickle.dump(model, open('m.pkl','wb'))`
Pickle load	`pickle.load(open('m.pkl','rb'))`
Joblib save	`joblib.dump(model, 'model.pkl')`
Joblib load	`joblib.load('model.pkl')`

Serialization enables model deployment at scale