Module 10 Summary: Capstone & Ethics

Data Science with Python – BSc Course

What We Covered

Module 10: Capstone & Ethics (L45–L48)

This final module brought together everything you learned across 48 lessons into integrated projects with ethical awareness.

Four Sessions:

- L45: Project Work 1 – Scope definition, data collection, initial EDA
- L46: Project Work 2 – Model training, evaluation, deployment start
- L47: ML Ethics – Bias detection, fairness metrics, regulatory compliance
- L48: Final Presentations – Live demos, peer review, Q&A

Core Theme: From technical skills to professional competence – building complete ML systems with ethical responsibility.

You went from zero to deployed data scientist in 10 modules

L45 Recap: Project Work 1

Problem Solved: How do you plan and start a data science project?

Key Takeaways:

- Define clear, scoped project problem with measurable success metric
- Select appropriate data sources (1000+ rows, 5+ features minimum)
- Create structured project plan with milestones and deliverables
- Set up professional folder structure (data/, notebooks/, src/, app/, models/)

Session 1 Deliverables:

- Problem statement (1–2 sentences)
- Data loaded and explored (shape, dtypes, missing values)
- At least 2 EDA visualizations (target distribution, correlations)
- Project repository initialized with requirements.txt

Good project structure makes debugging easier and impresses reviewers

L46 Recap: Project Work 2

Problem Solved: How do you build, evaluate, and deploy ML models?

Key Takeaways:

- Complete feature engineering (handle missing, encode categoricals, scale numericals)
- Train and compare at least 2 models with proper train/test split
- Evaluate with appropriate metrics (not just accuracy for imbalanced classes)
- Build deployment (FastAPI endpoint or Streamlit dashboard)

Critical Rules:

- No data leakage – fit scaler on training data ONLY
- Use sklearn Pipeline to prevent leakage automatically
- Model saved with joblib.dump() for deployment
- Practice demo at least twice before presentation

Working demo beats perfect slides – have a backup plan (screenshots)

L47 Recap: ML Ethics

Problem Solved: How do you detect and mitigate bias in ML models?

Key Takeaways:

- Bias enters at data, feature, label, and evaluation stages
- Measure fairness across demographic groups (80% rule for disparate impact)
- Use SHAP for model explainability (required by regulators)
- Monitor for bias before AND after deployment

Regulatory Reality:

- US ECOA requires explanation for credit denial
- EU GDPR Article 22 mandates right to explanation
- EU AI Act classifies credit scoring as high-risk
- Build ethical practices into workflow, not as afterthought

Fair does not mean equal outcomes – it means equal opportunity

L48 Recap: Final Presentations

Problem Solved: How do you communicate technical work to stakeholders?

Presentation Structure (5 minutes):

- Problem & Motivation (30 sec) – What and why?
- Data & EDA (1 min) – Source, size, key insights
- Model & Results (1.5 min) – What you tried, metrics, comparison
- Live Demo (1.5 min) – Show API or dashboard in action
- Lessons Learned (30 sec) – What surprised you? What would you change?

Grading Criteria:

- Clarity: Can a non-expert understand? (5 pts)
- Technical depth: Did you demonstrate ML understanding? (5 pts)
- Demo quality: Does deployment work? (5 pts)

Good Q&A demonstrates understanding better than any slide

Key Concepts Mastered

Professional Workflow

- End-to-end project execution from problem definition to deployment
- Code quality: reproducible, documented, version-controlled
- Evaluation: appropriate metrics, confusion matrices, residual plots
- Communication: translating technical work for non-technical audiences

Ethical ML

- Algorithmic bias detection and mitigation strategies
- Fairness metrics (disparate impact, demographic parity)
- Model explainability (SHAP values, feature importance)
- Regulatory compliance (ECOA, GDPR, EU AI Act)

Technical excellence without ethical awareness is incomplete

Finance Applications

Fair Lending & Credit Scoring

- Audit credit models for bias across demographic groups
- Implement 80% rule to detect disparate impact
- Use SHAP to explain individual credit decisions
- Document model decisions for regulatory review

Real-World Examples

- Apple Card gender bias investigation (2019)
- Amazon AI recruiting tool that penalized women
- Goldman Sachs algorithmic credit limits scrutiny
- Fed SR 11-7 model risk management guidance

Professional Practice:

- Monitor deployed models for fairness drift over time
- Maintain human-in-the-loop for high-stakes decisions

Responsible AI is legally required, not optional

Tools & Code Patterns

Project Structure Best Practices

- Folders: data/, notebooks/, src/, app/, models/
- Files: requirements.txt, README.md with setup instructions
- Version control: Commit early, commit often
- No hardcoded paths – use pathlib.Path

Key Libraries

- **joblib:** Model serialization for deployment
- **fairlearn:** Fairness constraints and mitigation
- **SHAP:** Model explainability for individuals
- **FastAPI:** REST API endpoints for predictions
- **Streamlit:** Interactive dashboards for demos

Code Quality:

- Functions with docstrings, random_state=42 everywhere
- sklearn Pipeline to prevent data leakage

Treat projects like production code – future employers will see it

Course Complete: Your Journey

From Zero to Data Scientist in 48 Lessons

What You Accomplished:

- Module 1–2: Python fundamentals and data manipulation (pandas, NumPy)
- Module 3: Statistics and visualization (hypothesis tests, matplotlib, seaborn)
- Module 4–5: Supervised learning (regression, classification, evaluation)
- Module 6: Unsupervised learning (clustering, PCA, pipelines)
- Module 7–8: Deep learning and NLP (neural networks, sentiment analysis)
- Module 9: Deployment (FastAPI, Streamlit, model serialization)
- Module 10: Integration and ethics (capstone projects, fairness)

You built:

- Real ML models trained on finance data
- Deployed applications with working demos
- Portfolio projects you can show employers

48 lessons, one integrated skill set – you are data scientists now

Course Complete: What's Next?

Continue Your Learning Journey

Deepen Your Skills:

- Kaggle competitions (practice on real problems with leaderboards)
- Fast.ai courses (practical deep learning, free online)
- Andrew Ng's ML Specialization (Coursera, foundational theory)

Build Your Portfolio:

- GitHub profile with 3–5 polished projects (README, clean code, demos)
- Blog posts explaining your approach (demonstrates communication)
- Contribute to open-source ML projects

Finance Career Paths:

- Quantitative analyst (algo trading, risk modeling)
- Data scientist in fintech (credit scoring, fraud detection)
- Bloomberg Terminal / financial data engineering
- ESG investing and responsible finance (CFA Institute certificates)

The best way to learn data science is to DO data science – keep building