Special Topics: Machine Learning (ML) for Networking

COL867

Holi, 2025

Lecture 1

Tarun Mangla

Welcome

- Introductions
- About this course
- What will we cover
- Setting the rules
- Attend the talk at 4p

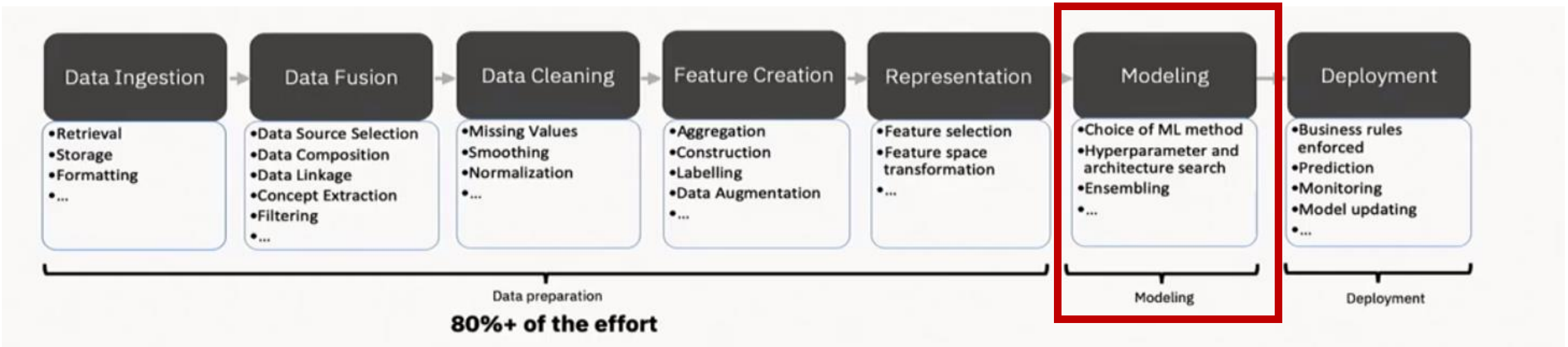
Why This Course

- Machine learning (ML) wave everywhere
- Internet is no exception!
- As a networking researcher/engineer, I want to understand how ML can be used to make networks more fast, secure, efficient ..
- Easy-peasy: Take an ML course, right?
- But ..

How is This Course Different From an ML Course

- Existing ML courses are geared towards ML researcher (or mathematician)
- Everything (or most) I want to know about ML is as a “user”
- This describes most people who are “doing ML”
- Discussion: What do you think “doing ML” is about?

The Machine Learning Pipeline



Data preparation:

- **Affects the final predictive accuracy** -- generally *more* than the modeling step does!
- Like modeling, also *contains parameters* which **should be tuned**
- **80%+ of the effort** in a data project is in data preparation (some say 90%+)
- Not treated in textbooks: left as black art → gives rise to **many conceptual errors** in practice -- *most* errors in data science happen in data preparation

Credit: Alex Gray

Learning Objectives

- **[Learning] problems in networking**
  - Identify different learning problems in networking that enhance network security, efficiency, and performance
  - Explore the role of ML in solving these learning problems

- **ML pipelines [in networking]**
  - Understand the stages of ML pipelines including data collection, data representation, model evaluation
  - Learning strategies to develop ML models that are robust, explainable, and performant
  - All within a networking context but can be applied to other domains

What will We Cover [Tentative]

Week	Topic
Week 1	Introduction + Network Warmup
Week 2	Application classification
Week 3	Performance inference
Week 4	Security
Week 5	Resource management
Week 6	Data representation
Week 7	Data collection
Week 8	Foundational Models
Week 9	Foundational Models 2
Week 10	Explainability
Week 11	Formal verification
Week 12	Robustness
Week 13	Synthetic data generation
Week 14	Project presentation

Course Material

- No textbook
- Cover a set of papers
- Resources and reading list will be up on the course page

Workload and Evaluation [Tentative]

- Exams – 40%
  - Minor – 20%
  - Major – 20%
- Assignments – 15%
- Project [in pairs] – 30%
- In-class participation [quiz, in-class exercises] – 15%
- **Audit Policy:** Need to score at least **B grade** for audit pass

Assignments

- 3 programming assignments – 5% each
- Goal: Hands-on experience (not super hard)
- **Late Policy**
  - You have late hours: 72hours
  - Don’t ask me for extensions – not fair to other students
  - (Rare) Exceptions in extenuating circumstances can be accommodated.

Project

- **Goal**: Research-oriented
- A set of problems will be shared. You are welcome to work on problem of your own (discuss with me)
- Work in groups of 2
- Project milestones
  - Proposal
  - Mid-term review
  - End-term evaluation

In-class participation

- **Goal:** Test your understanding of course material
- Done through:
  - Quiz
  - Hands-on exercise
- Questions to expect
  - Key concepts covered in the lectures
  - Relevant concepts from the reading list
- Expect 1 quiz or hands-on exercise every week
- Best 90% will count towards final grade

Logistics

- **Communication Channels**
  - Piazza for logistics and discussions
  - Assignments via Moodle
- **No Attendance Policy** – But regular attendance will make the course much easier
- **Office Hours**: Wednesday 3-4p, Bharti 423
- **Plagiarism**
  - Please read the IITD Honor code
  - Please acknowledge your collaborators, *specifically*
  - **Any cheating will get you an F**

Looking Ahead

- Upcoming week
  - What is ML for networks
  - Hands on networks data exploration
- Do this in advance!
  - Join Piazza
  - Install jupyter notebook
- Please reach out if you have any questions!

Next

Title: Leveraging LLMs for Networking & Security in Cloud Environments
Speaker: Deepak Bansal, Microsoft
Date/Time/Venue: January 3, 2025/4:00 pm - 5:00 pm/Bharti 501

Abstract: Customer networks have grown, mostly organically, large and complex in cloud environments like Azure. Customers are often afraid to make changes and find it hard to diagnose when things go wrong. In this talk, I am going to share how Microsoft is using LLMs to simplify network operations at scale in Azure and how it is enabling the same for its customers through Azure Copilot. On the security side, I will share how LLMs are being used to enable security monitoring and threat hunting.

Bio: Deepak graduated from IIT D in CS in 1999 and did a Master's in CS at MIT. He is currently a Corp Vice President and Technical Fellow at Microsoft in Redmond, WA USA and is driving cloud (Azure infrastructure) and security (Microsoft's Secure Future Initiative).