Open Source MLOps: Course Intro

Alex Kim @alex000kim

Course Instructor: Alex Kim

- Physics -> Software Engineering -> Data Science -> MLOps
- Worked on ML projects involving satellite image and video data, log analytics, mining sensor data, etc.
- Developed and taught DS & ML curriculums for EdTech bootcamps and universities
- Some facts about me:
  - Live in Montreal, 🇨🇦
  - Dad to two toddlers
  - Enjoy baking
- @alex000kim on GitHub, Twitter and LinkedIn
- https://alex000kim.com

What's MLOps?

Best practices, standards and tools for ML application development and deployment

- Dataset Management & Versioning
- Model Development
- ML Experiment Tracking
- CI/CD for model re-training
- Model Monitoring
- Team collaboration
- Progress reporting
- …

source: https://en.wikipedia.org/wiki/MLOps
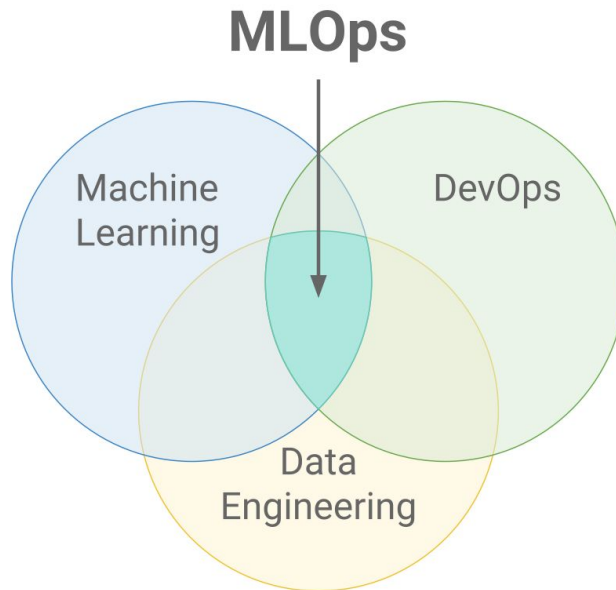
About the Course

- Focus on "forest behind the trees"
- Real end-to-end project
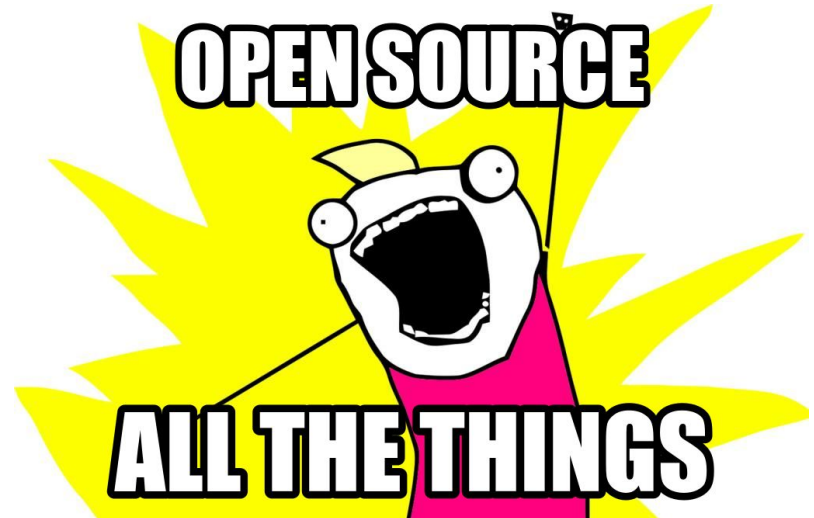- Detailed step-by-step instructions
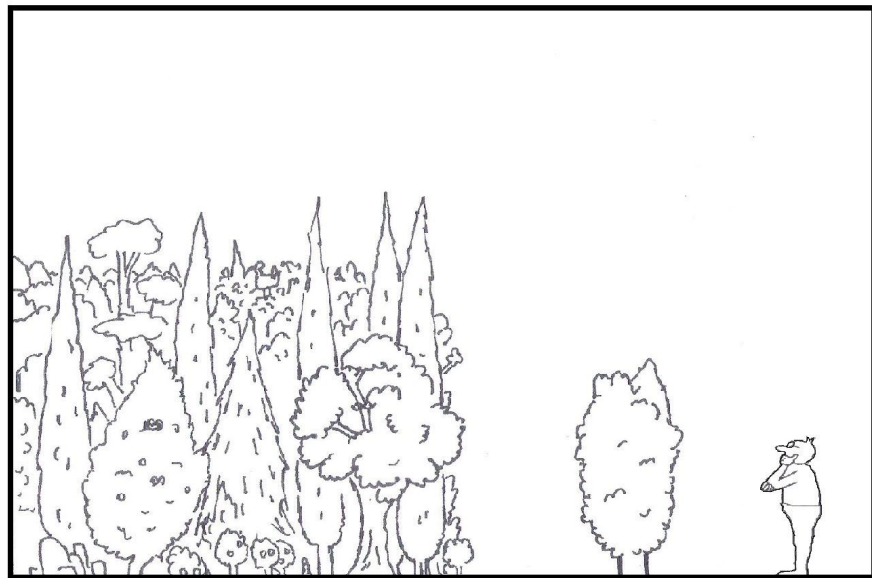
P. KLAMMER

Open-source MLOps

When in doubt, go with Open-Source Software:

- Community and support
- Low/zero cost
- Public roadmap and discussions

Week 1: Kick-Starting an ML Project

- Apply best practices for establishing ML project structure and dependencies management
- Manage project dependencies with pip and virtualenv
- Version datasets with DVC

Week 2:
ML Pipelines, Reproducibility and Experimentation

- Refactor a Jupyter notebook into a reproducible ML pipeline
- Version artifacts of an ML pipeline in a remote storage
- Iterate over a large number of ML experiments in a disciplined way
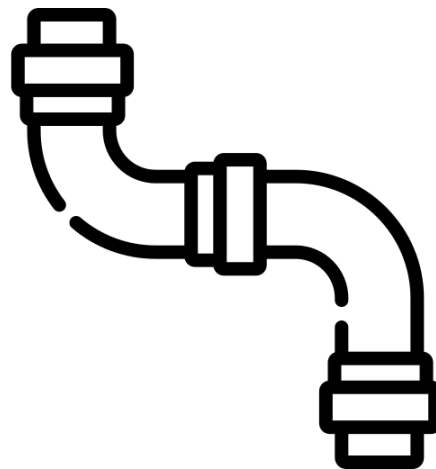
Week 3: CI/CD for ML and ML-based Web API

- Learn the basics of CI/CD
- Leverage the power of CI/CD tools for ML projects with CML
- Integrate an ML model into the FastAPI framework
- Build and test a Docker container running a web API service
- Deploy the resulting Docker container to cloud
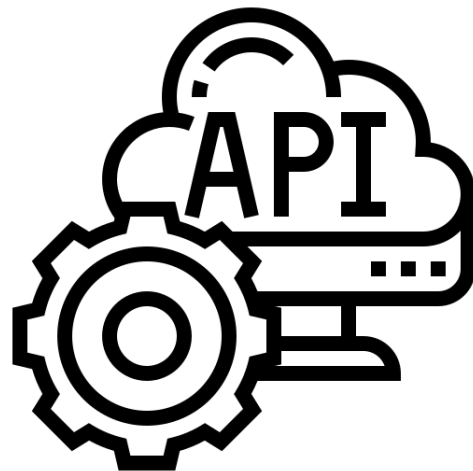
Week 4:
Data Drift Monitoring for ML Projects

- Distinguish between application monitoring and ML monitoring
- Use Alibi Detect framework to detect data drift

What's your experience with ML and MLOps?