

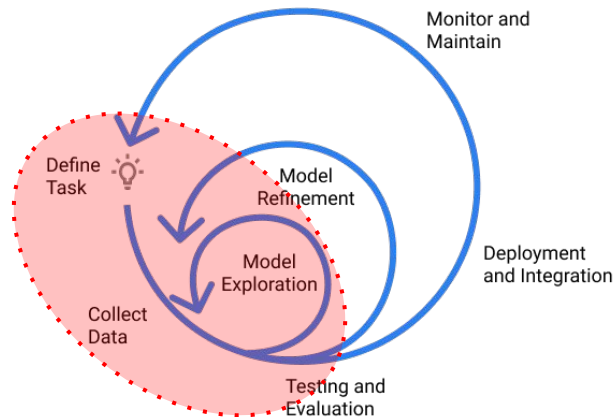
ML project lifecycle and MLOps

Alex Kim @alex000kim 

ML project lifecycle: proof of concept (POC) phase

- Goal:
Prove to stakeholders that the solution is feasible
- Activities:
Iterate between task definition, data discovery/collection and model performance

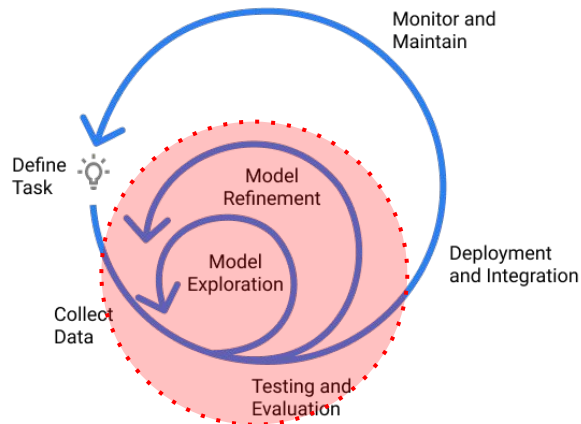
Machine Learning Development Lifecycle



ML project lifecycle: Model Refinement

- Goal:
Introduce structure and rigor to ML experimentation (aka reproducibility)
- Activities:
Code refactoring, ML pipeline development, experiment management

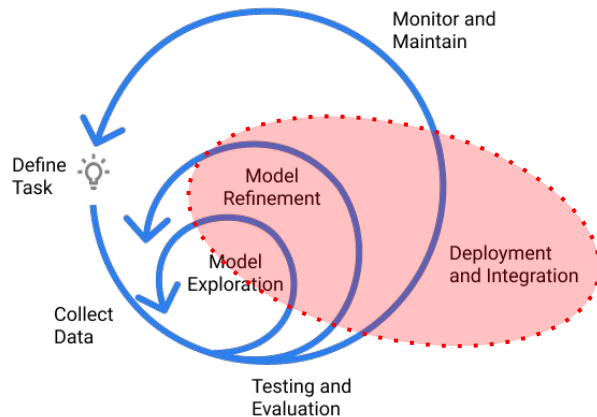
Machine Learning Development Lifecycle



ML project lifecycle: Model Integration

- Goal:
Integrate ML model into larger application
- Activities:
Integration with other app components,
e2e testing, CI/CD pipelines for model
retraining and performance validation

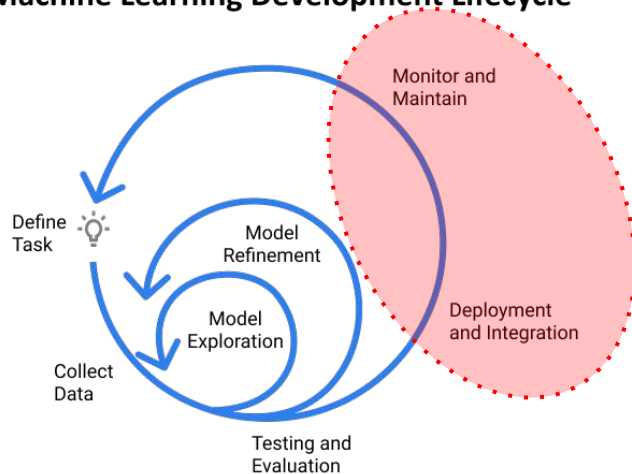
Machine Learning Development Lifecycle



ML project lifecycle: Deployment and Monitoring

- Goal:
Automated deployment and ML monitoring
- Activities:
CI/CD pipelines for model deployment,
data drift monitoring

Machine Learning Development Lifecycle



ML Project Structure

Why is it important?

- Standardization
- Code quality
- Collaboration
- Reproducibility

“Because that default project structure is logical and reasonably standard across most projects, it is much easier for somebody who has never seen a particular project to figure out where they would find the various moving parts.”

- [Cookiecutter Data Science](#)

ML Project Structure

```
$ tree -L 2
```

```
.
├── Dockerfile
├── LICENSE
├── README.md
├── requirements.txt
├── data
│   ├── processed
│   └── raw
├── metrics.json
├── models
│   └── clf-model.joblib
├── notebooks
│   └── TrainChurnModel.ipynb
├── params.yaml
├── reports
│   ├── feat_imp.csv
│   └── figures
├── src
│   ├── app
│   ├── stages
│   └── utils
```

DVC for Data Versioning



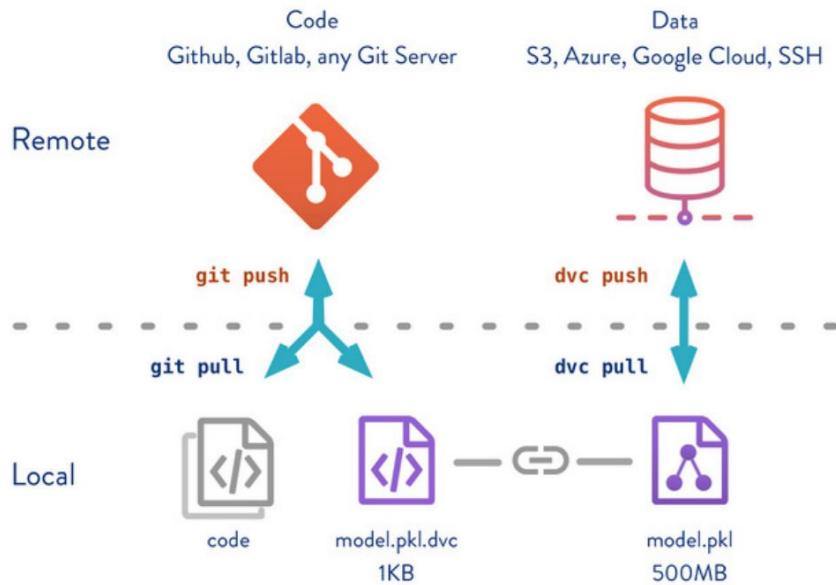
What's DVC?

- Command-line tool:
 - Data Versioning
 - ML Pipelines*
 - Experiment Management*

*with integrated data versioning

- Open-source
- Integrates with Git
- Language- platform- storage-agnostic

DVC for Data Versioning



Practice time!

