Deployment of Machine Learning Models

Section 3.3 Notes

Links

The Google + Microsoft papers referenced in the lecture:

1. "The ML Test Score: A Rubric for ML Production Readiness and Technical Debt Reduction" (2017) Breck *et al*. *IEEE International Conference on Big Data* (Google)

Download URL: https://research.google/pubs/pub46555/

2. "Software Engineering for Machine Learning: A Case Study" (2019) Amershi *et al*. (Microsoft)

Download URL: https://www.microsoft.com/en-us/research/uploads/prod/2019/03/amershi-icse-2019_Software_Engineering_for_Machine_Learning.pdf

Shadow Deployments: https://christophergs.com/machine%20learning/2019/03/30/deploying-machine-learning-applications-in-shadow-mode/

Monitoring ML models: https://christophergs.com/machine%20learning/2020/03/14/how-to-monitor-machine-learning-models/

The below resources are on more advanced topics that we will not be covering in the course

- Google's *Site Reliability Engineering* is one of the best references out there, it's available for free here: https://landing.google.com/sre/sre-book/toc/index.html

Deployment of Machine Learning Models

- Martin Fowler's testing guide is pretty comprehensive. If you are new to testing this may be overwhelming:: https://www.martinfowler.com/testing/

- *Obey the Testing Goat* by Harry Percival is a good applied introduction to Test Driven Development (TDD): https://www.obeythetestinggoat.com/

Advanced, narrow vs. broad integration tests:
https://martinfowler.com/bliki/IntegrationTest.html