ML lifecycle

There are many flavours

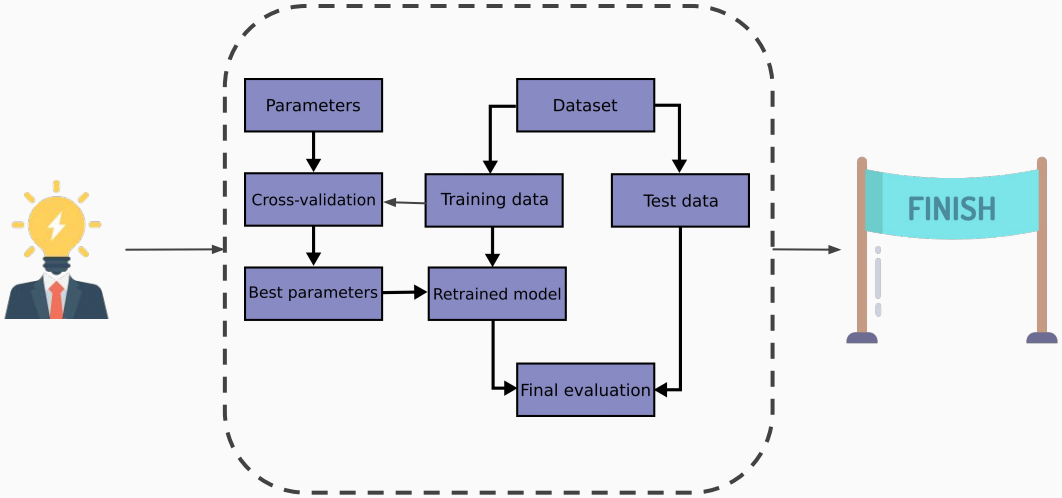
Naive view of ML

Naive thoughts

1. I have a great idea to apply Machine Learning.
2. I will just train a model and then see the error.
3. My job is done.

Naive thoughts

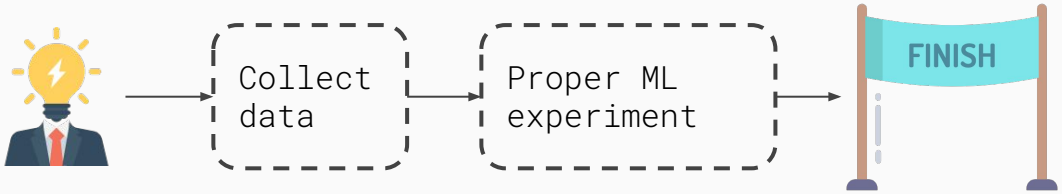
1. I have a great idea to apply Machine Learning.
2. I will make a proper Machine Learning experiment.
3. My job is done.



Adapted from: https://scikit-learn.org/stable/modules/cross_validation.html

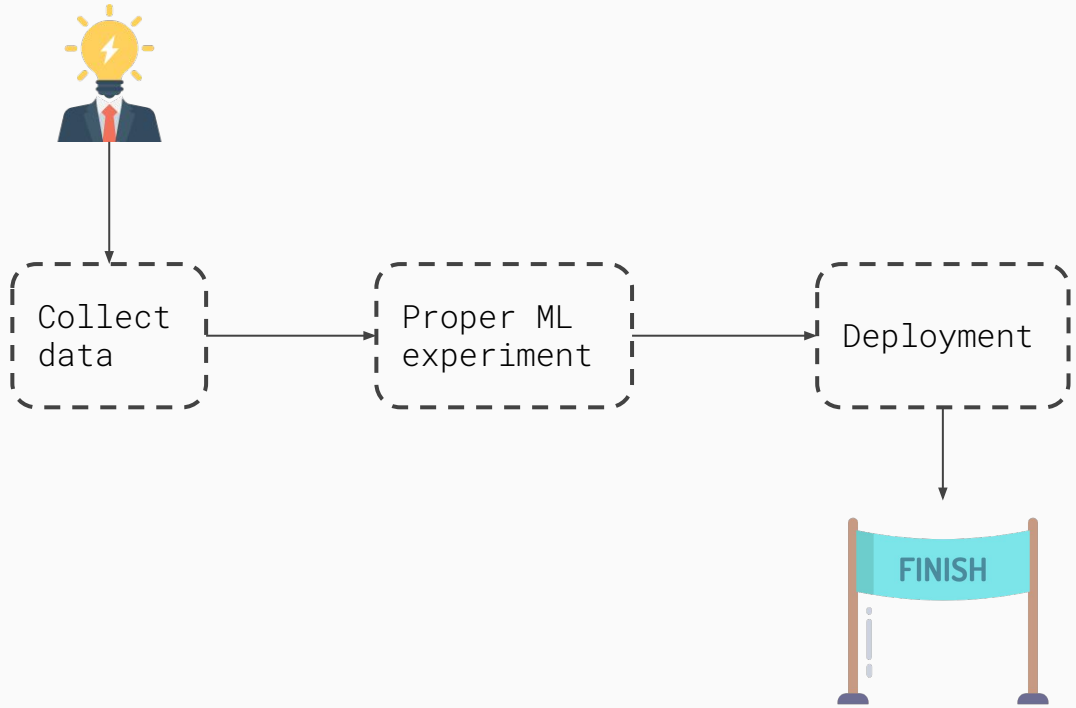
Naive thoughts

1. I have a great idea to apply Machine Learning.
2. I will collect the data I need
3. I will make a proper Machine Learning experiment.
4. My job is done.

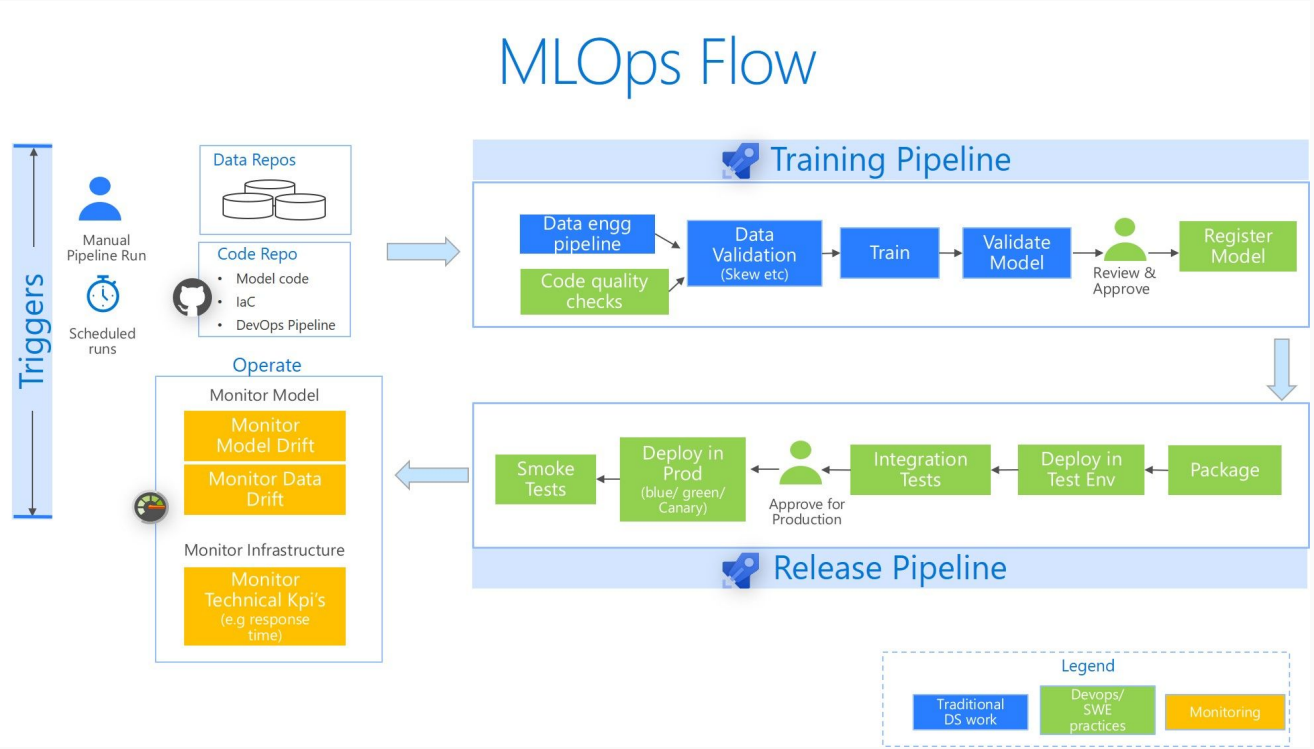


Naive thoughts

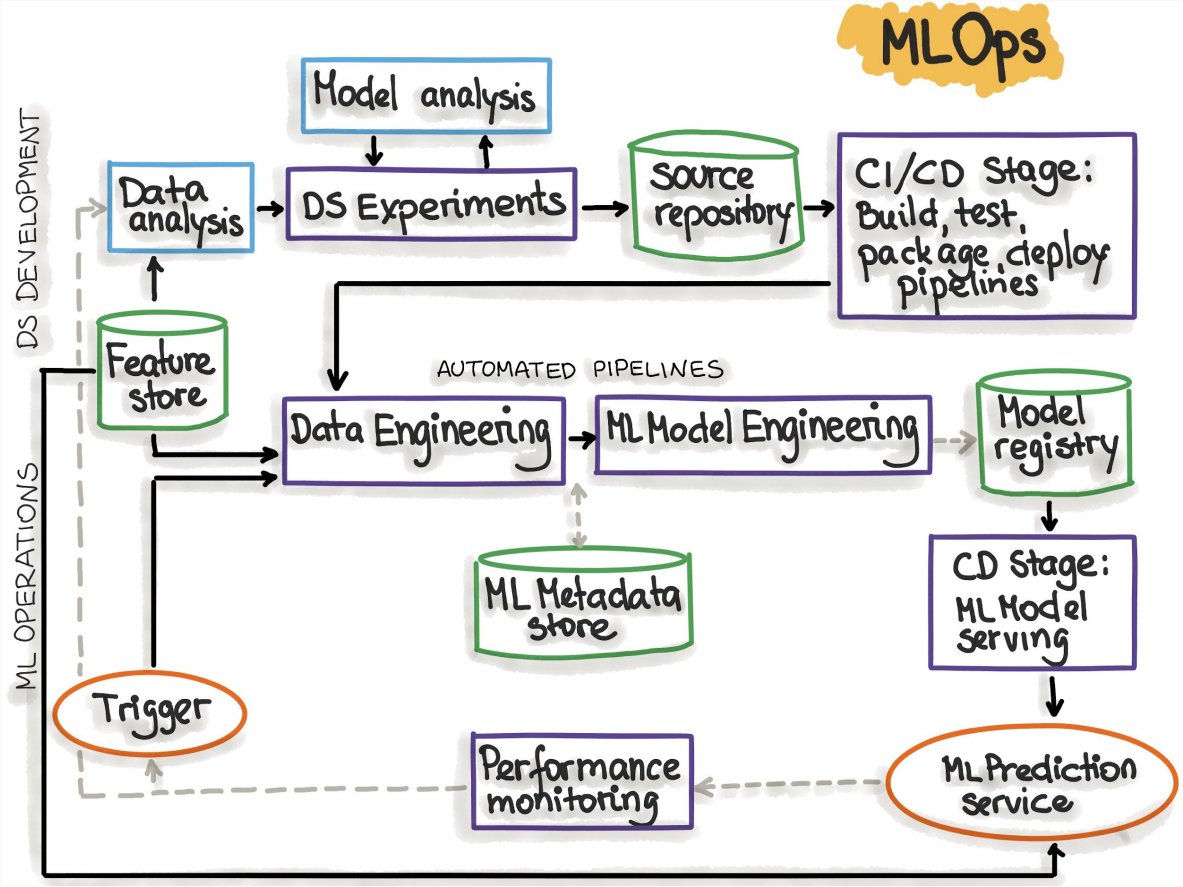
1. I have a great idea to apply Machine Learning.
2. I will collect the data I need
3. I will make a proper Machine Learning experiment.
4. I will deploy my model.
5. My job is done.



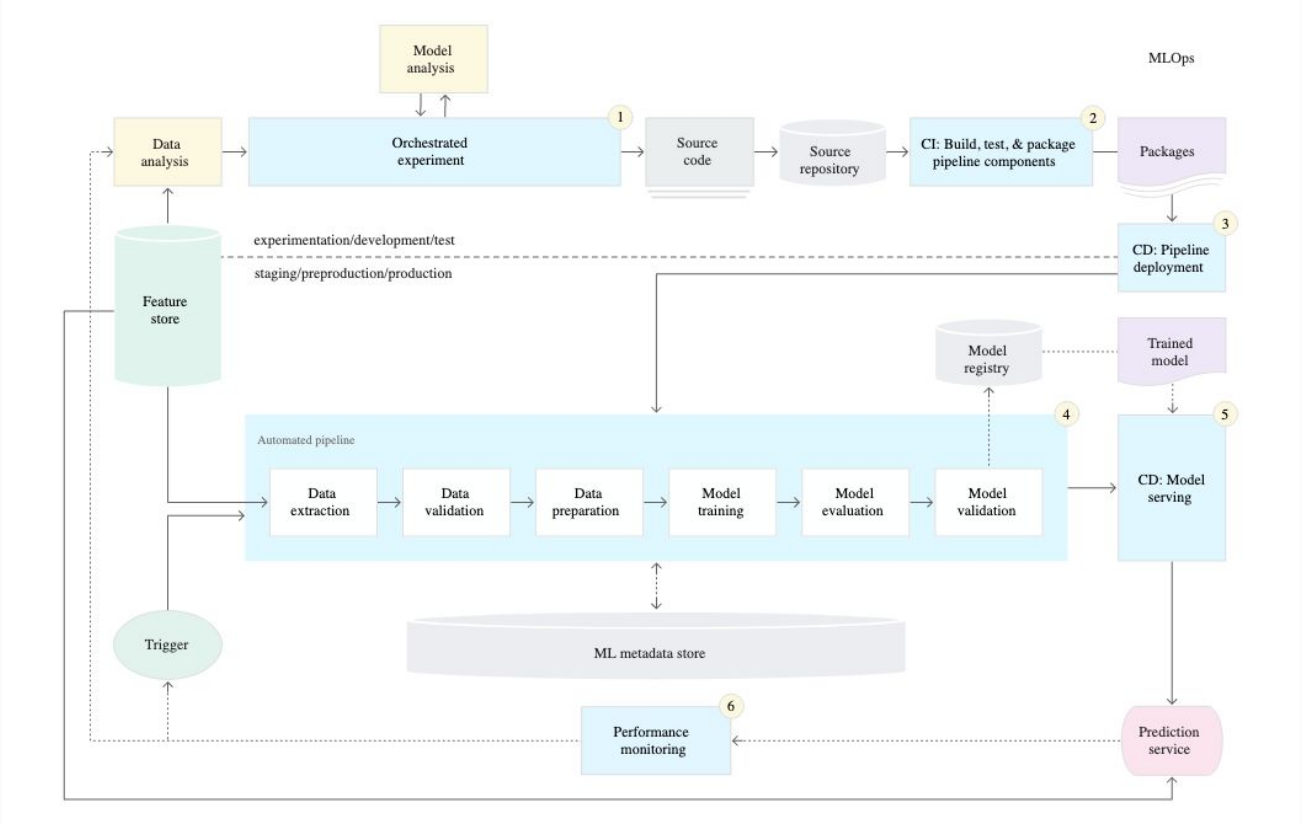
MLOps



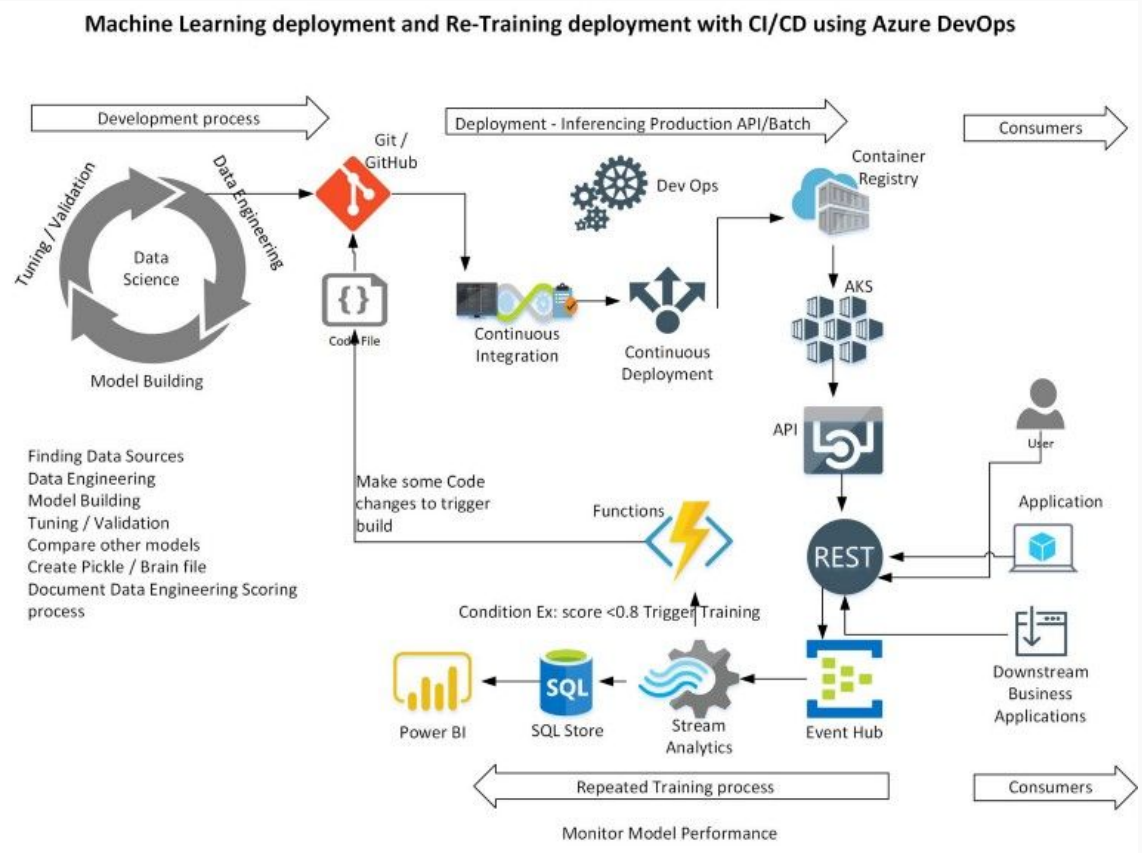
https://github.com/rsethur/MLOps



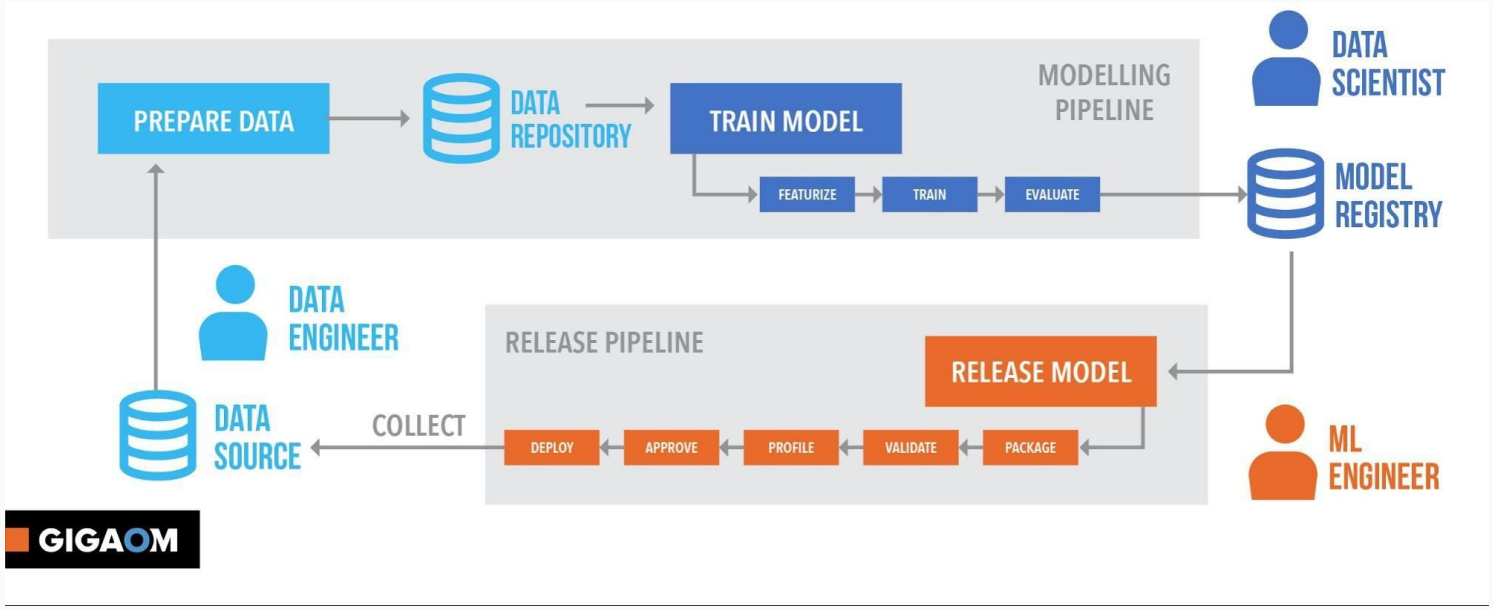
https://ml-ops.org/content/mlops-principles



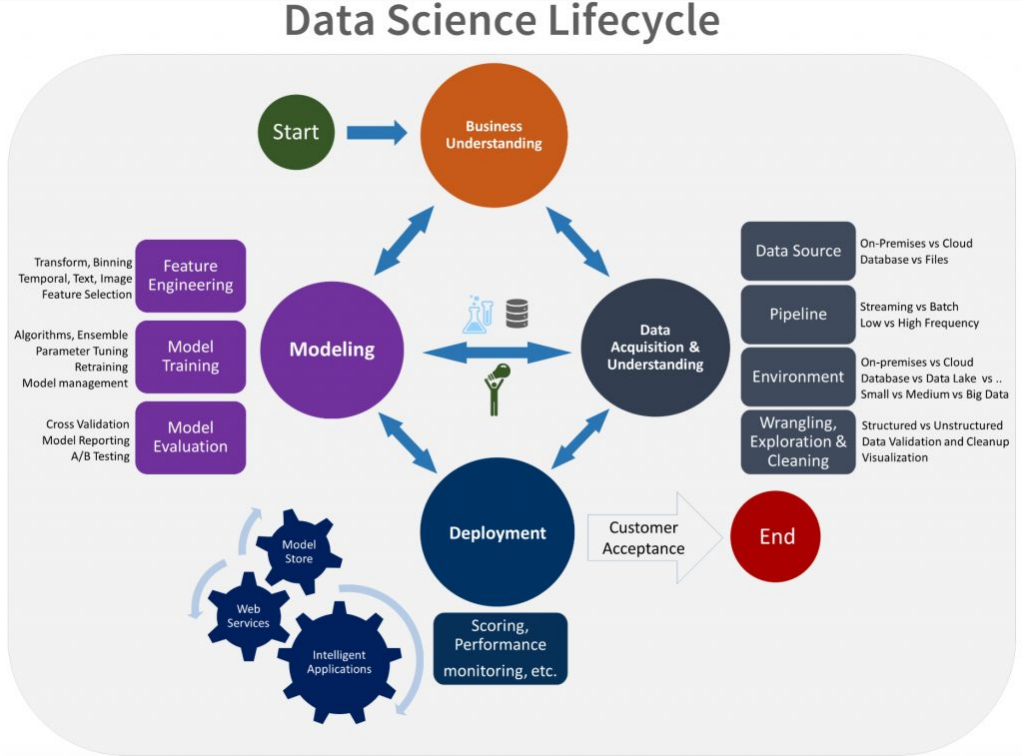
https://cloud.google.com/solutions/machine-learning/mlops-continuous-delivery-and-automation-pipelines-in-machine-learning



https://medium.com/analytics-vidhya/azure-machine-learning-services-mlops-c0fde346a8ff

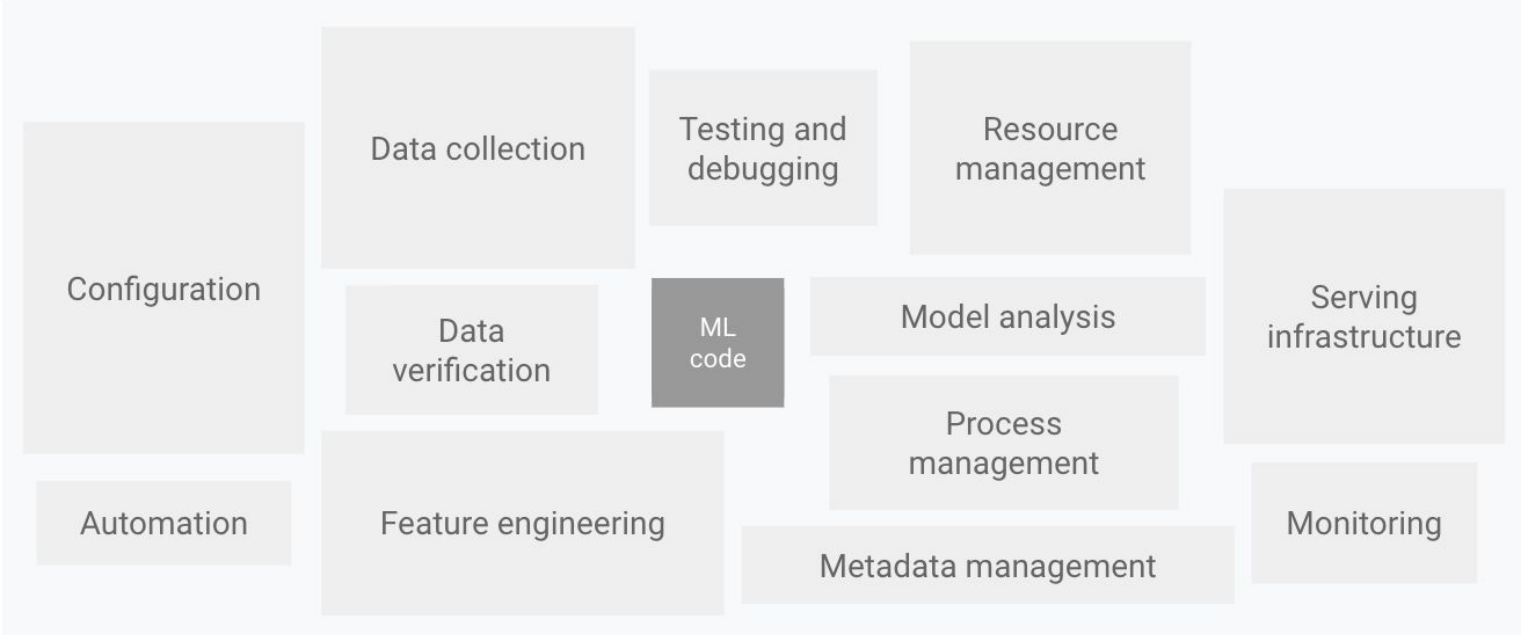


https://gigaom.com/report/delivering-on-the-vision-of-mlops/



https://www.run.ai/guides/machine-learning-operations/

Highlights



Everyone (roughly) agrees on this

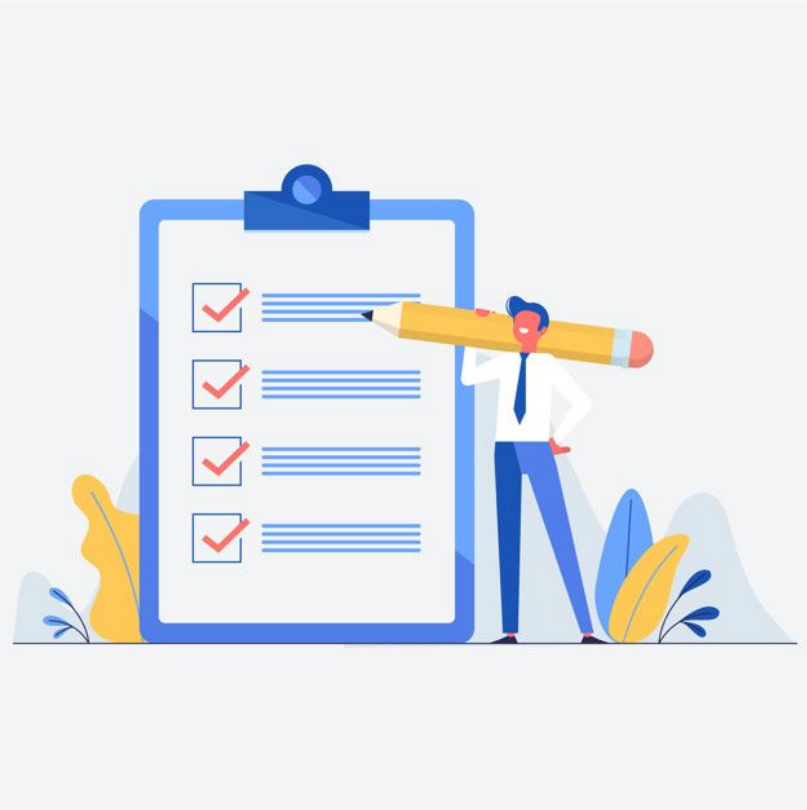
https://papers.nips.cc/paper/2015/file/86df7dcfd896fcaf2674f757a2463eba-Paper.pdf

Data Collection

- How often do we have new data points?
- How many sources do we have?
- Where is the data coming from ?
- How do we integrate all data sources?

Data Verification

- Is data consistent?
- Does it give you the information you need?
- Does it have the kind of values it is supposed to have?



Model Analysis

- What kind of errors is the model making?
- Is it biased towards some groups of people?
- How are the errors with respect to some relevant features in the dataset?
- Which features are more important?

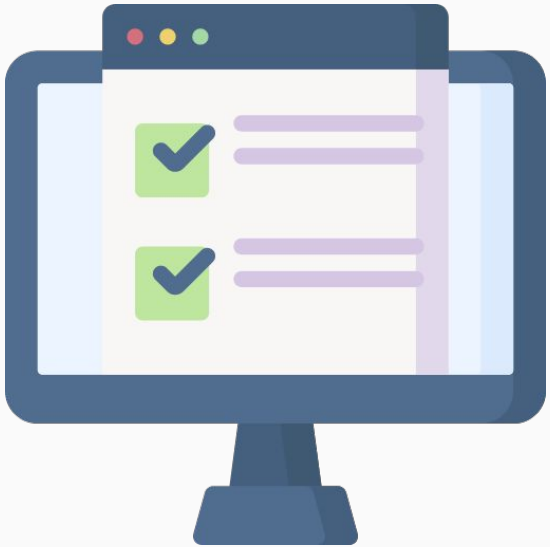
Monitoring

- How are the values we are getting for the model in production?
- Is performance drifting over time?
- Is the data distribution changing?

Testing and debugging

- Have you created an object that makes custom transformations?
- Have you created a new metric function?
- Have you created a new loss function?

Make sure your code does what you think it does!



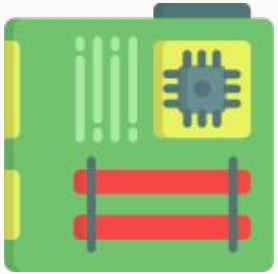
Serving

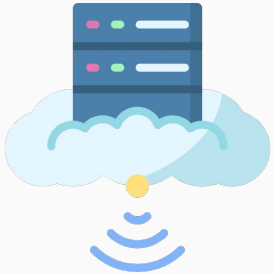
- How long does a request take?
- From which parts of the world is my model being called?
- How many times per second is my model being accessed?
- How has access to my model?
- What kind of computations is my model doing?
- How do I make sure that what runs on my computer runs somewhere else?

Everything can run on a CPU and everything has a one, while GPUs are supported by specialized algorithms and devices

Serving

- How long does a request take?
- From which parts of the world is my model being called?
- How many times per second is my model being accessed?
- How has access to my model?
- What kind of computations is my model doing?
- How do I make sure that what runs on my computer runs somewhere else?





Go for cloud most of the time, unless there are very specific security, scalability and money requirements

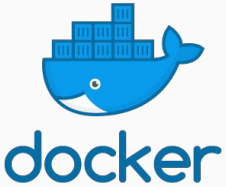
Serving

- How long does a request take?
- From which parts of the world is my model being called?
- How many times per second is my model being accessed?
- How has access to my model?
- What kind of computations is my model doing?
- How do I make sure that what runs on my computer runs somewhere else?

They are roughly the same thing...this decision is usually made at the company level, including aspects that go beyond a Machine Learning, but affect other areas

Serving

- How long does a request take?
- From which parts of the world is my model being called?
- How many times per second is my model being accessed?
- How has access to my model?
- What kind of computations is my model doing?
- How do I make sure that what runs on my computer runs somewhere else?



This is most likely the right approach most of the time, only use something else if you have to.