

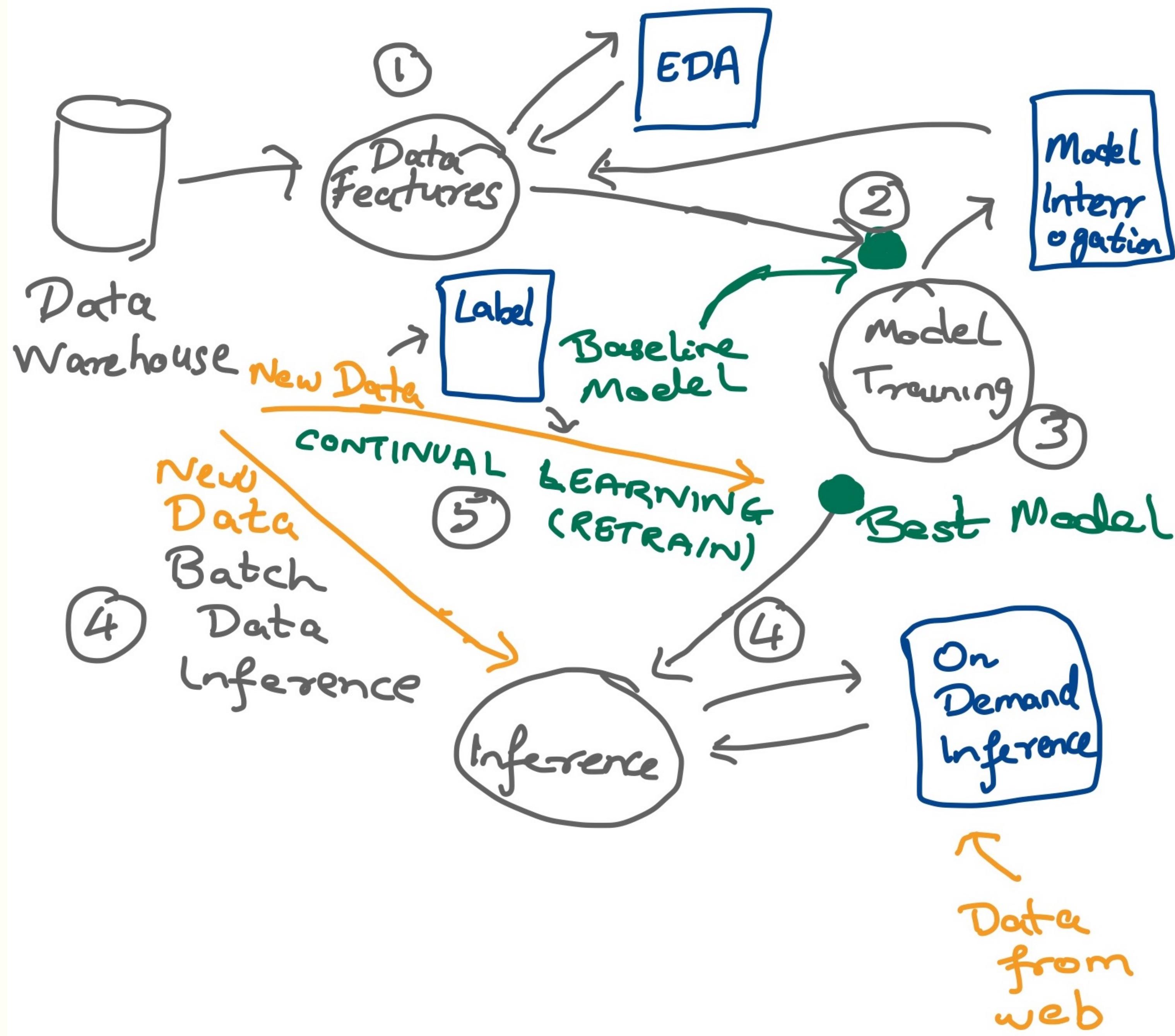
Zero To MLOps

A story in 3 acts

Rahul Dave: @rahuldave, @rahuldave@sigmoid.social

Acts

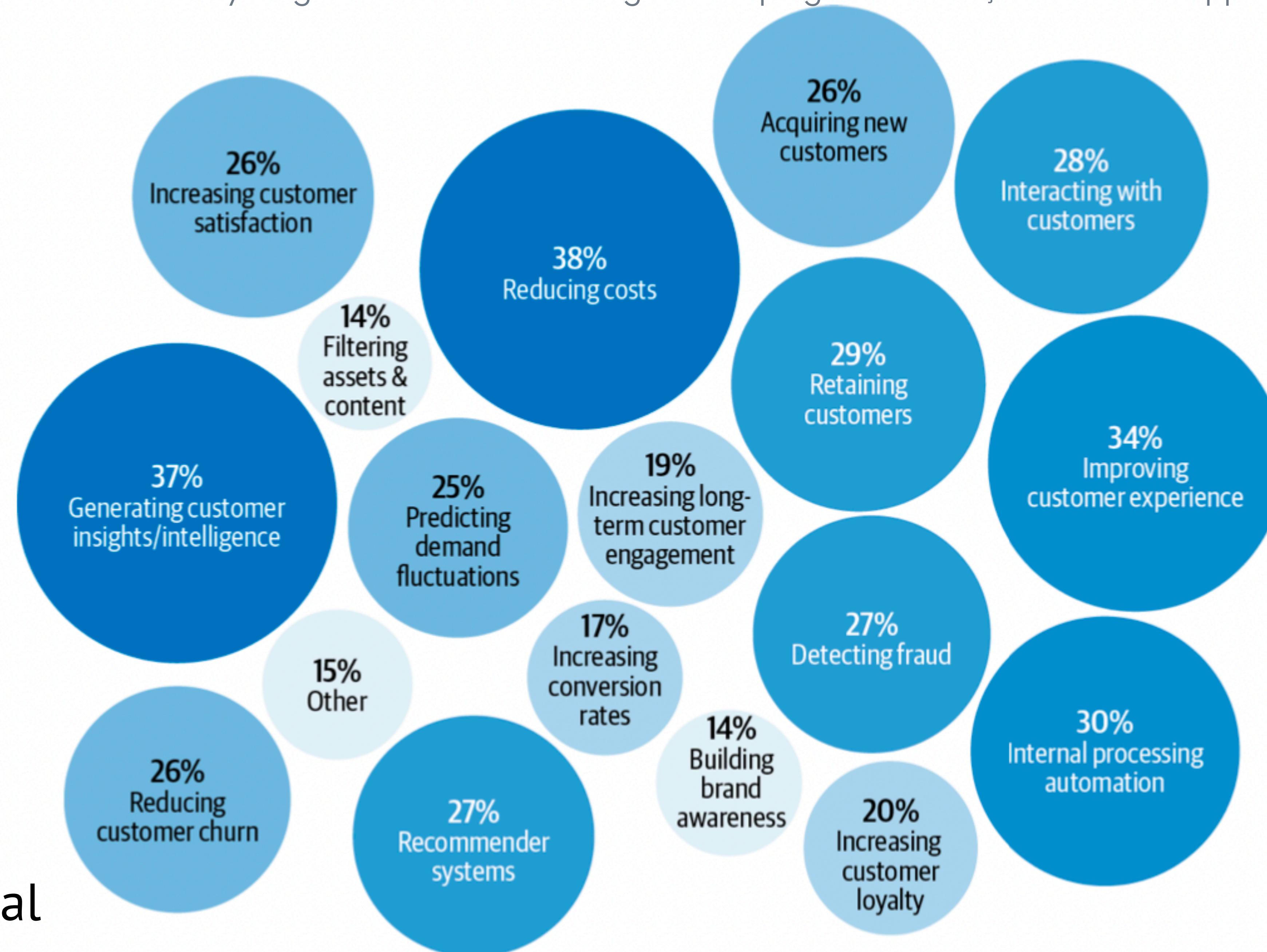
- Understanding the problem and creating re-usable code. Managing data and artifacts.
- Using re-usable code automatically and creating pipelines. Using pipelines in experiments to iterate through the data creation and modeling process.
Registering best models. Creating re-producible pipelines for use in inference.
- Monitoring performance. Handling multiple models and re-training and replacing models. Continual, Active, Learning to reduce the amount of labeling and to combat drift.



The Data Science Process we are creating infrastructure and a process for.

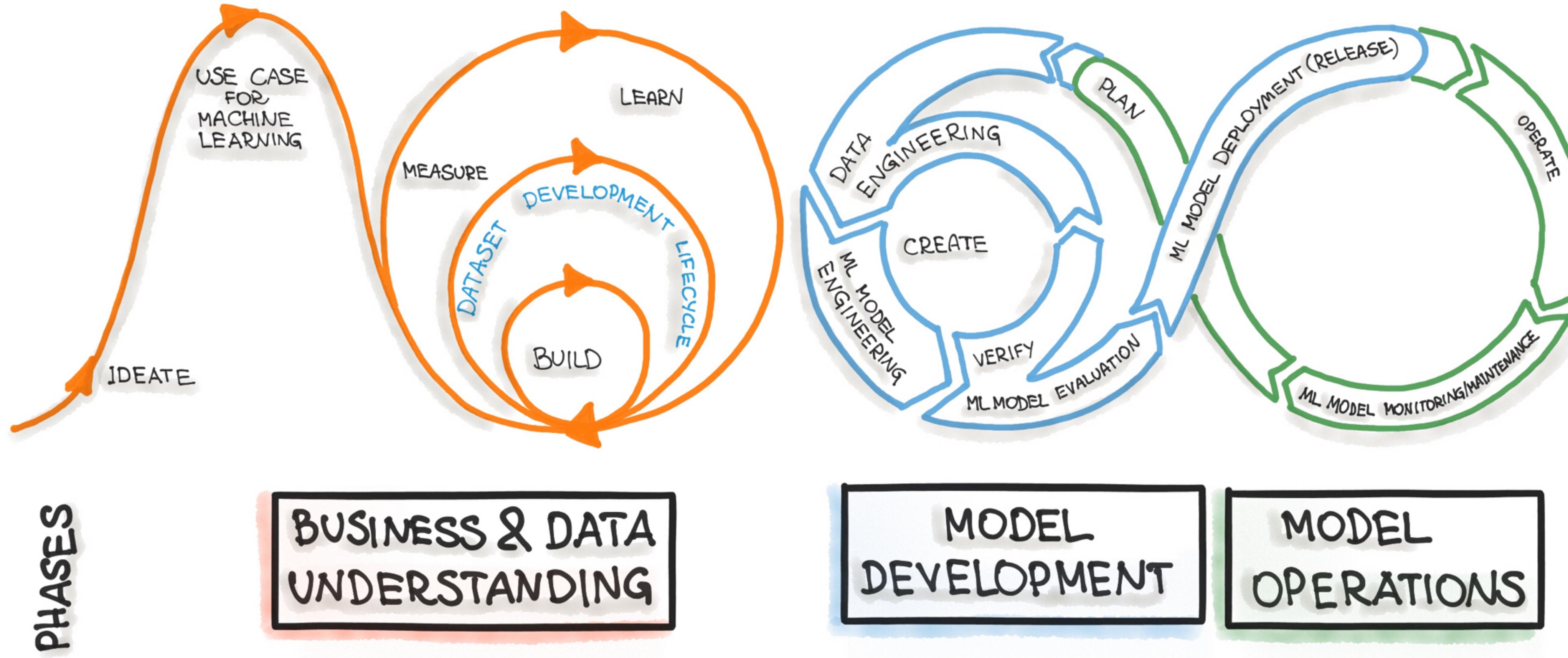
- (1) Data from the warehouse is processed into relevant features, on which
- (2) a baseline model is trained. If performance is decent, we
- (3) train better models.
- (4) The best performing model is used to make predictions (inference) on new data
- (5) Sometimes we want to train the model again when new data comes in, but this requires labeling carefully chosen parts (based on inference from the best performing model) of the data, and then retraining.

Most ML projects seem to be currently targeted at understanding and helping customers, and internal applications that increase efficiency.



from Huyen Et al

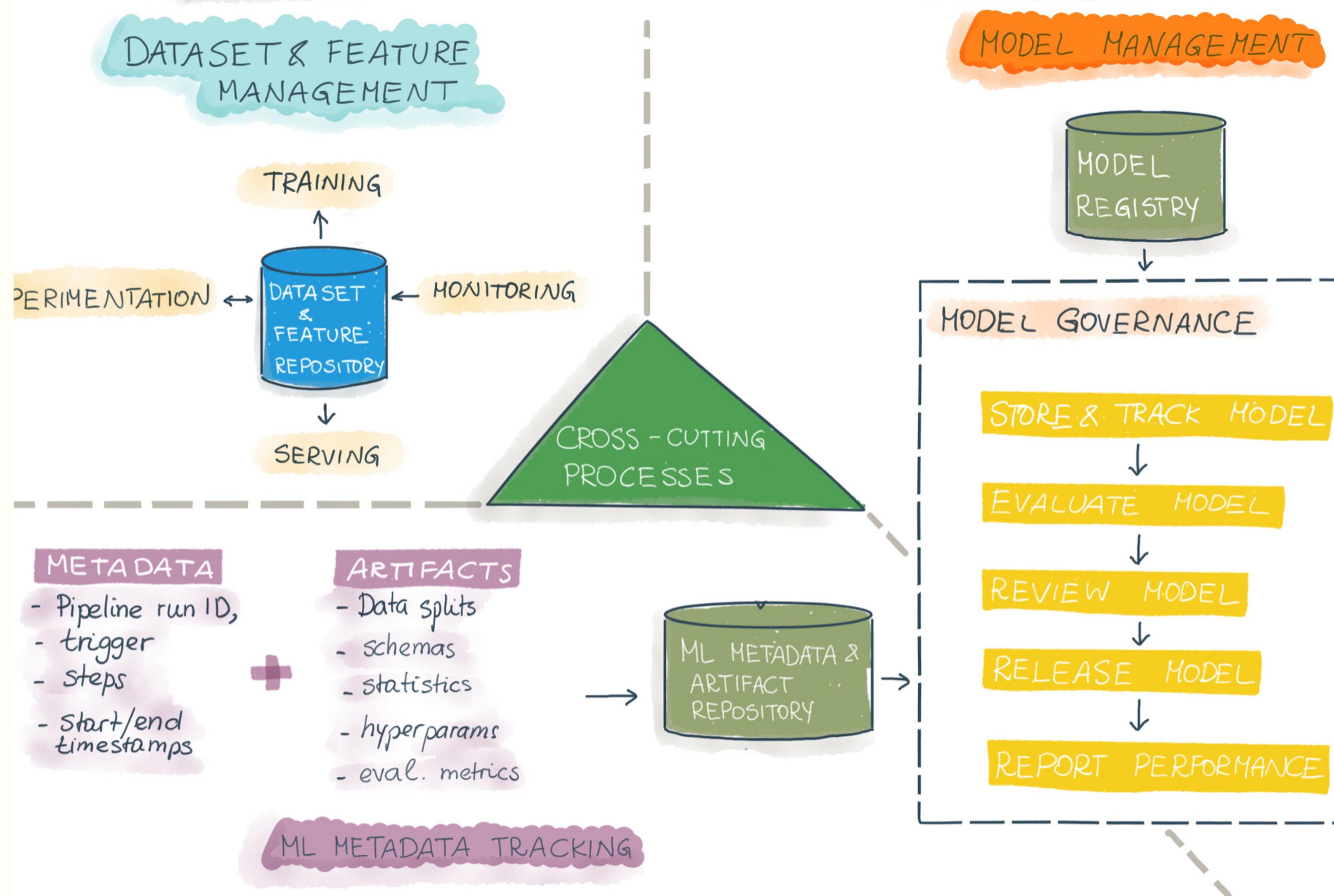
CRISP-ML(Q)



@visenger

The Machine Learning Cycle. After collecting the data and establishing our use case and business KPIs we do some data engineering. We then create a baseline model, and do further model engineering to beat the performance of the baseline model. We validate our models and choose the best performing models for inference. The model performance will decrease over time as new data comes in which is different from old data and so we re-train the model with (possibly carefully chosen and labeled parts of) the new data

DATA & MODEL MANAGEMENT



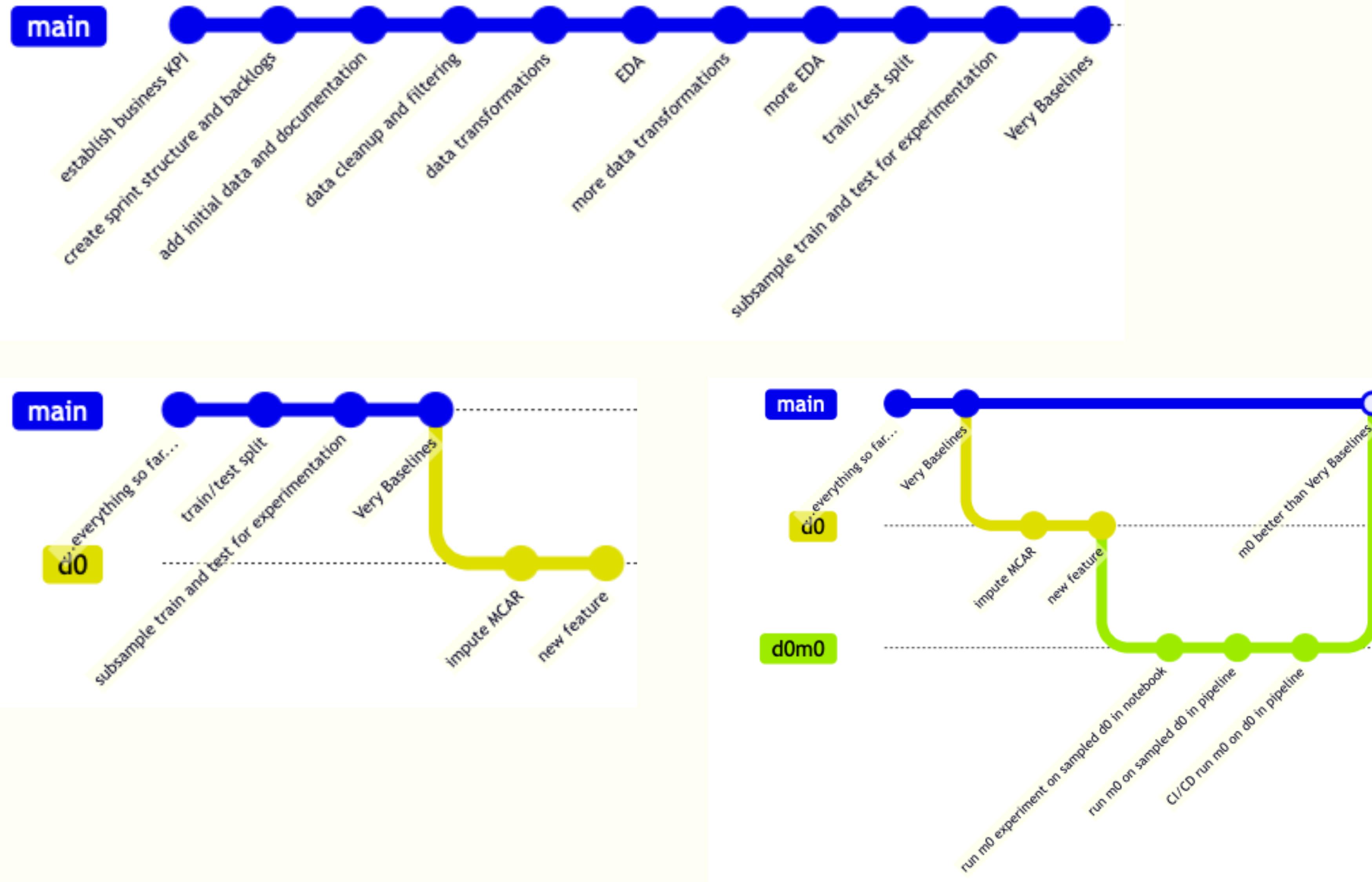
Every project has many considerations involving code, data, artifacts, and models.

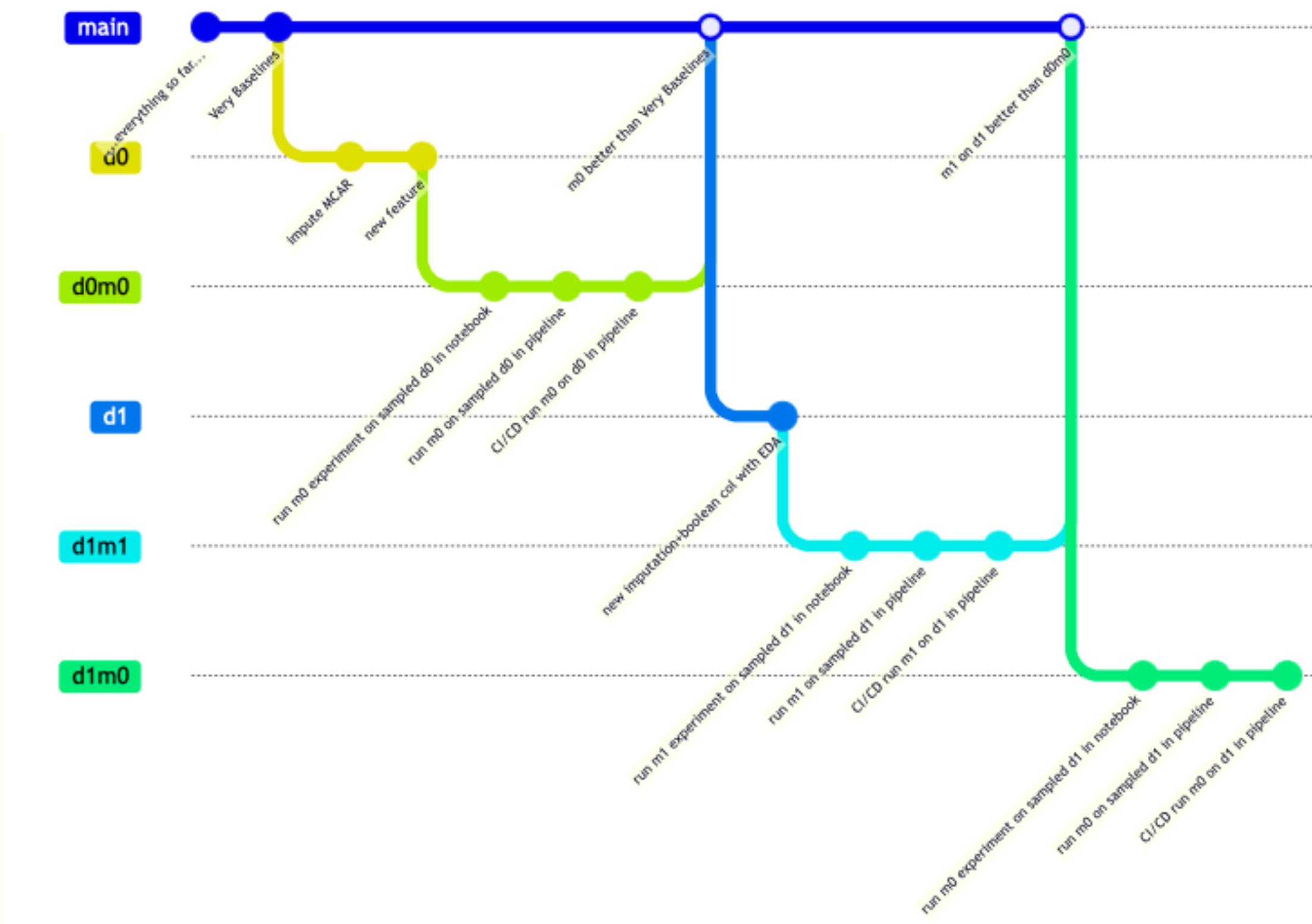
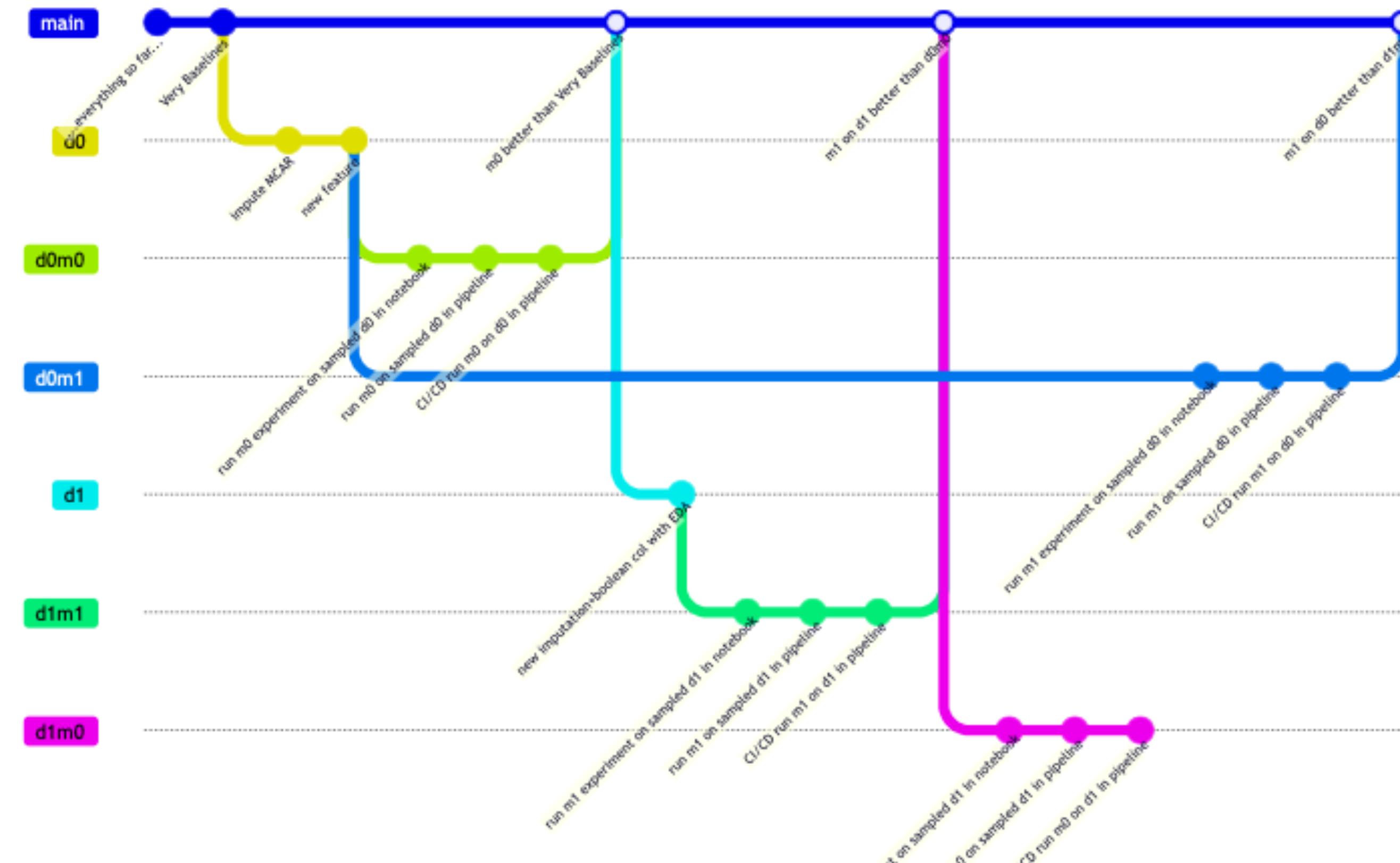
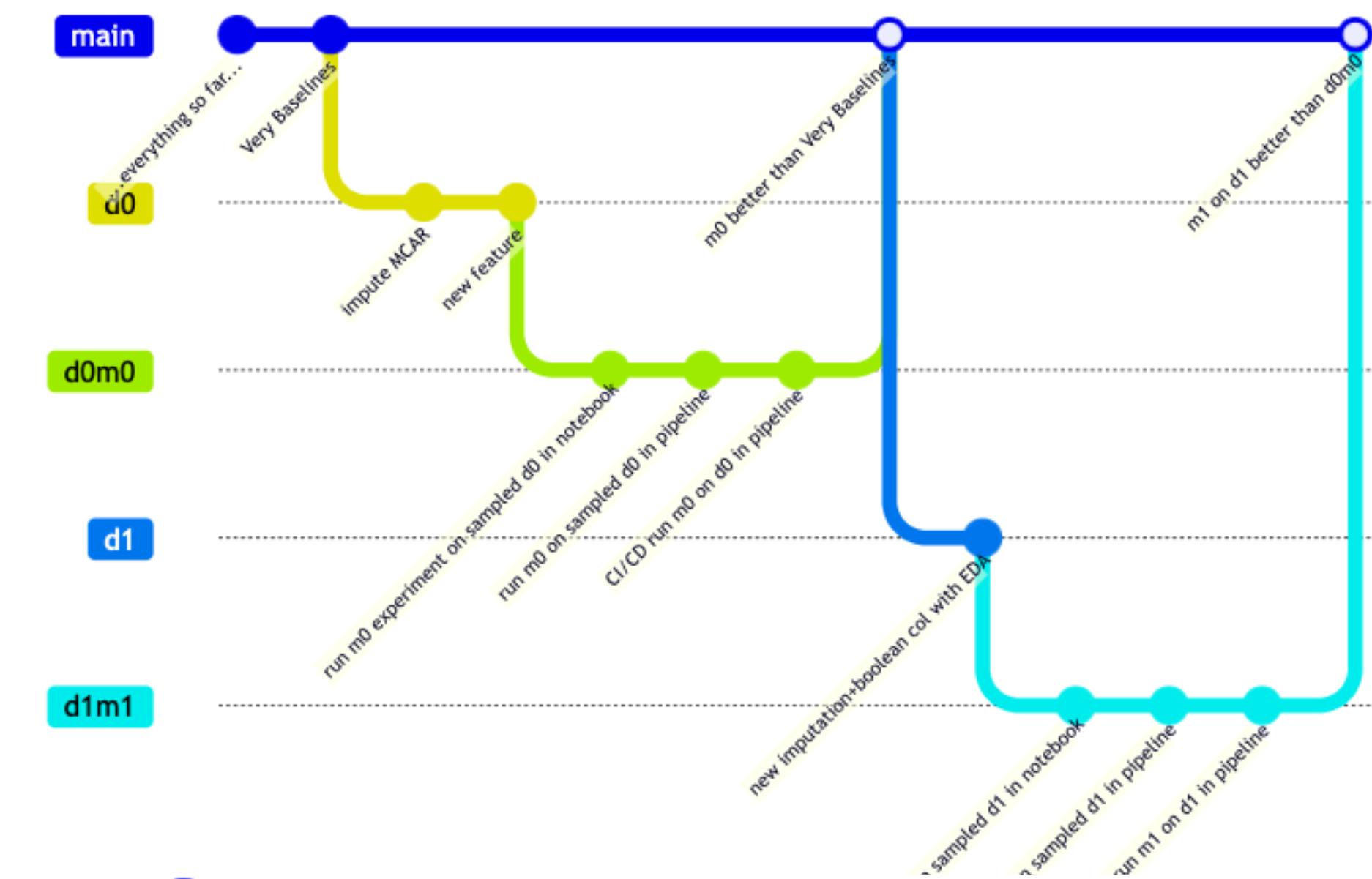
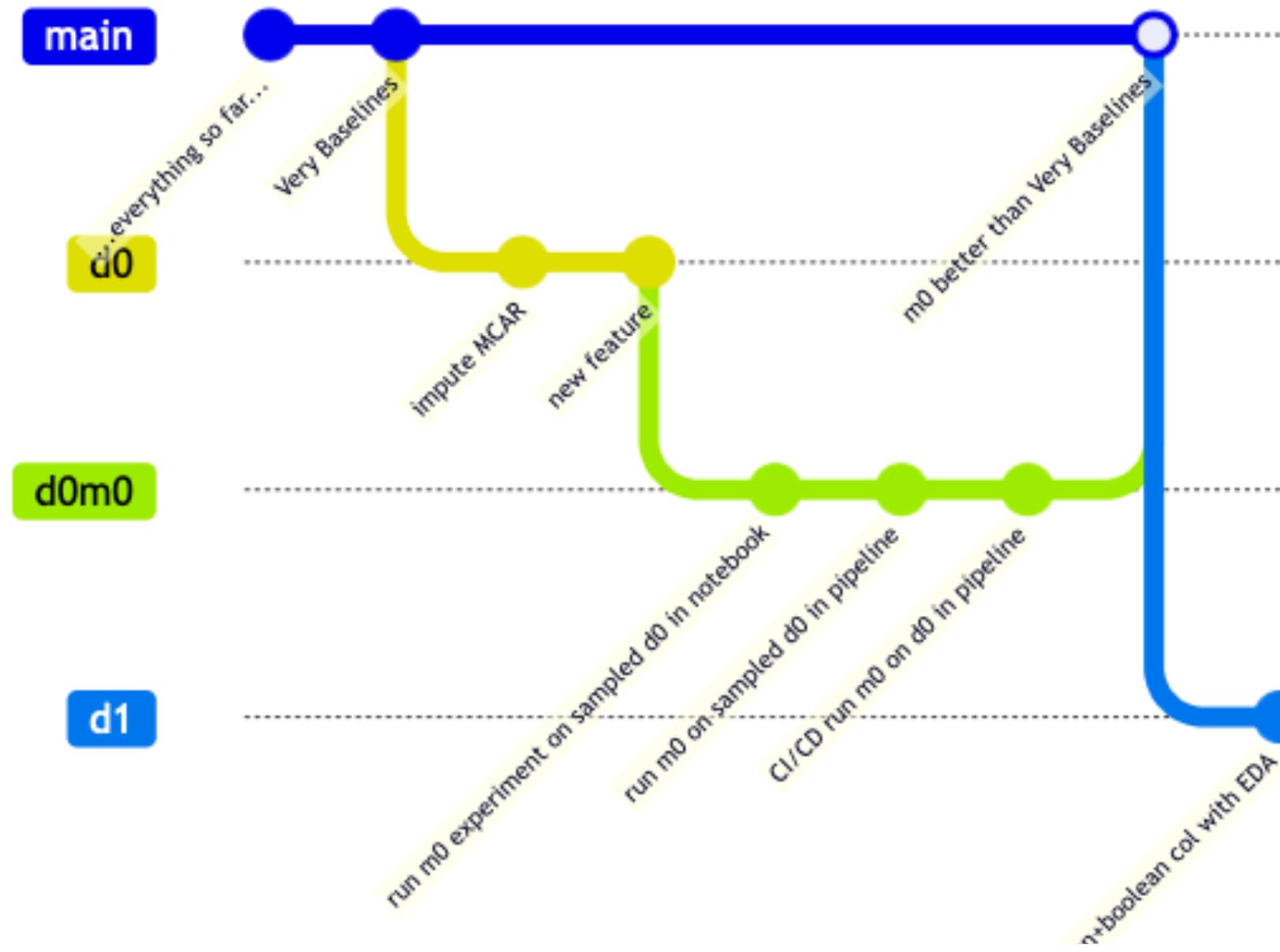
Needs of a ML Project

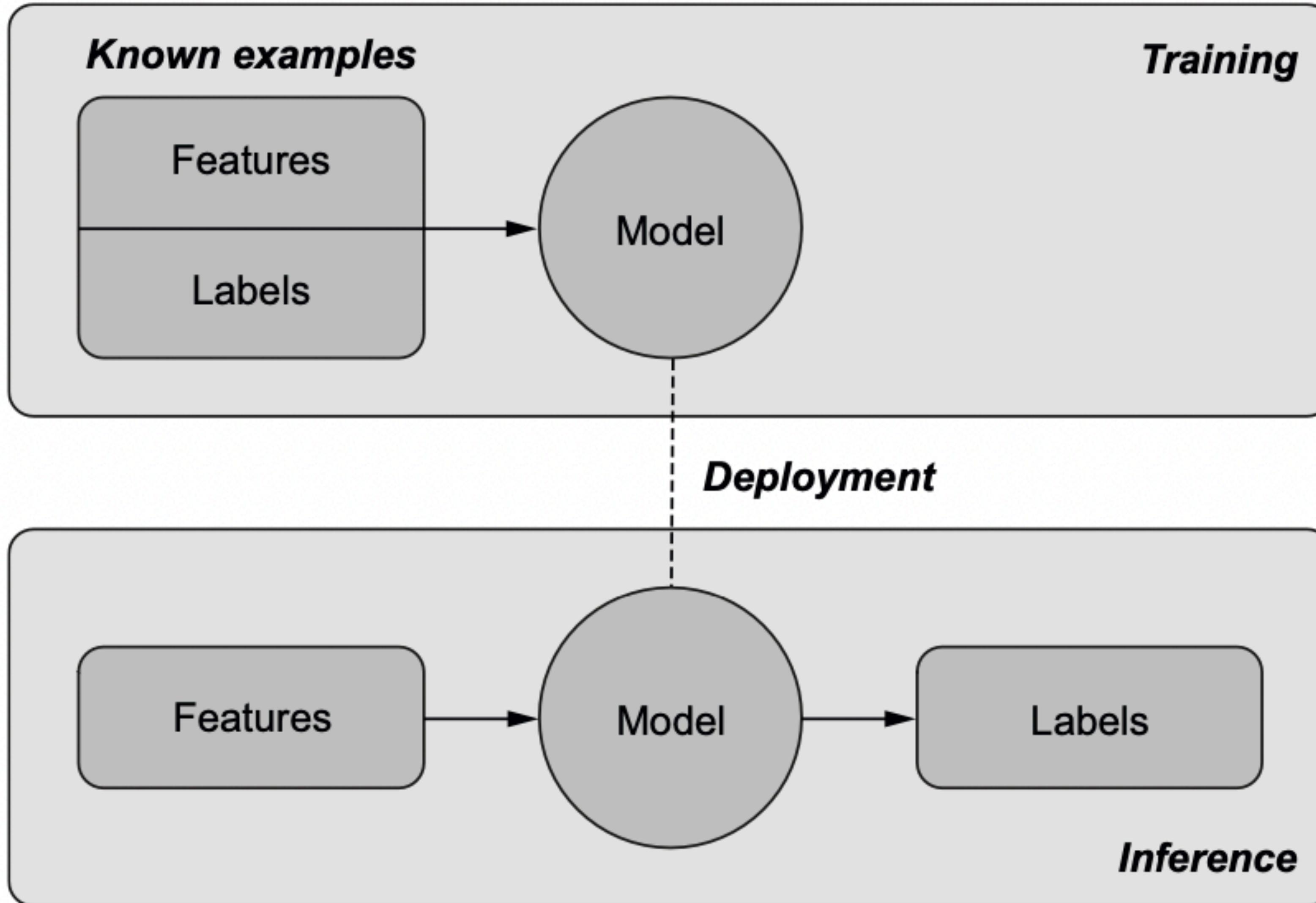
Three keystone variables govern success for a ML project, as seen in Shankar et al. (2022):
Shankar S., Garcia R., Hellerstein J. M., Parameswaran A. G. (2022) [Operationalizing machine learning: An interview study](#). arXiv.

1. Velocity: the speed at which we can iterate on models, thus being able to establish baseline performance early on, and hence decide which direction to take in a project or abort it all-together.
2. Validation: the process by which we make sure our data and results from any stage of our ML pipeline and from any model are sensible and correct
3. Versioning: our ability to recall any artifact with associated any one of our models or features for further deployment in downstream pipelines or production.

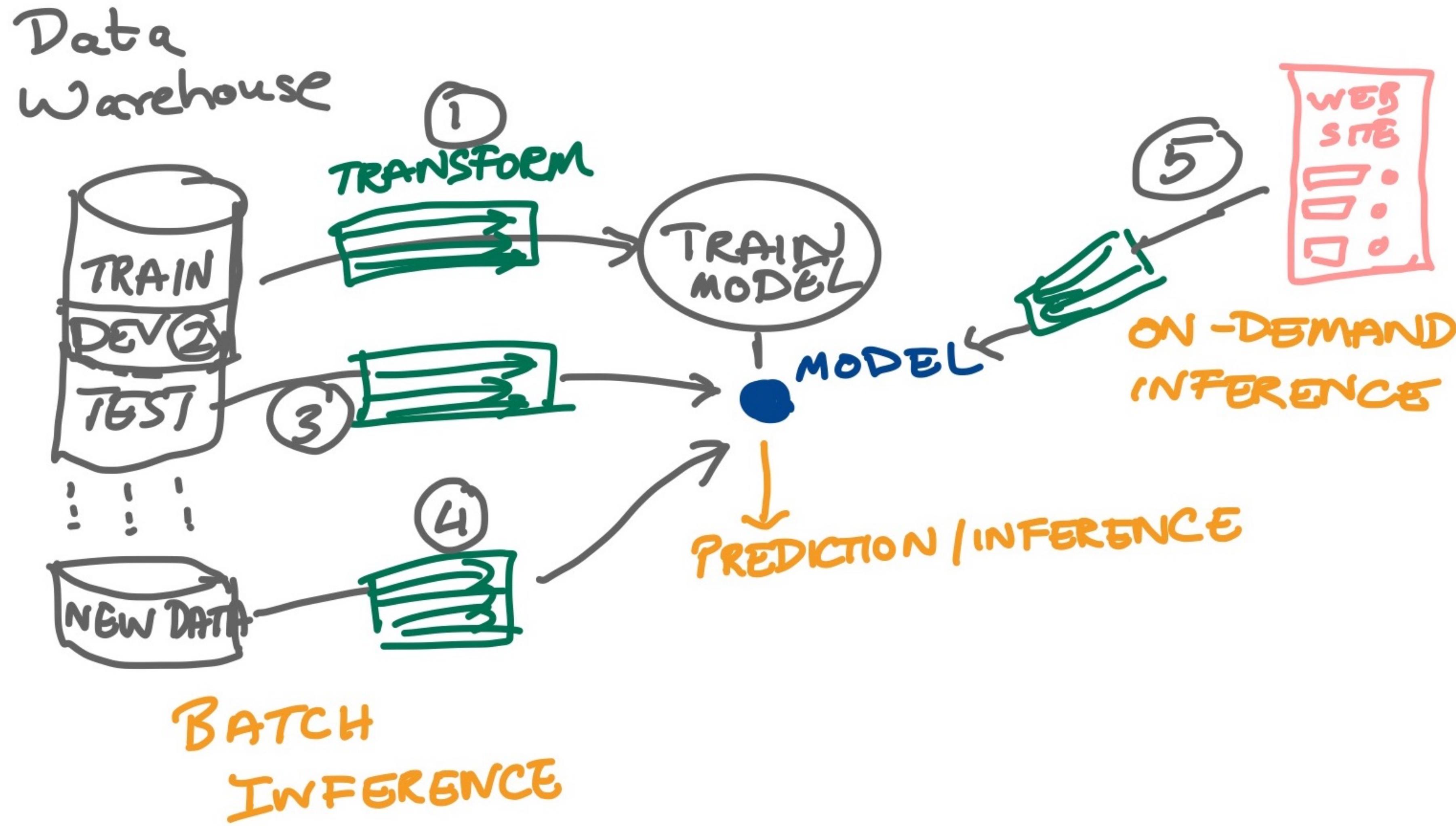
Project 1: Training a Classifier to do inference





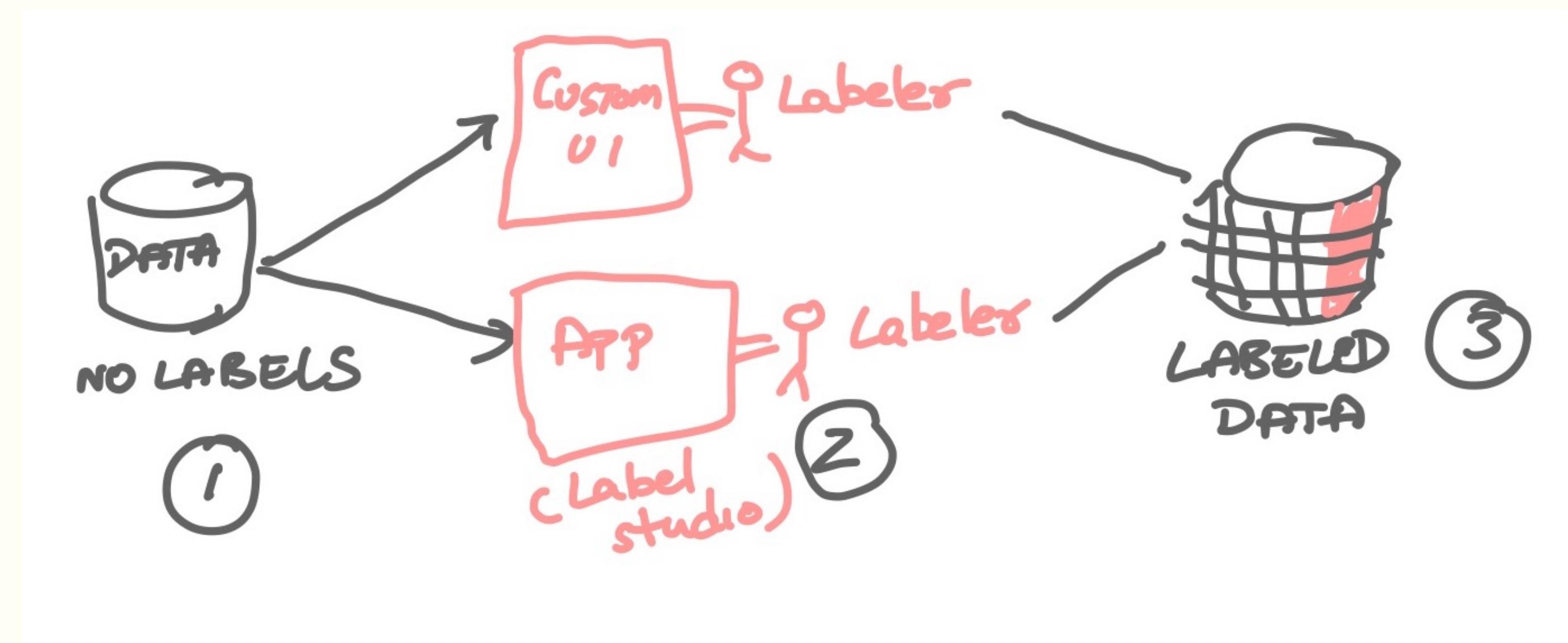


A model is trained using data and target labels. The model can now make predictions. The process of making predictions is called inference. Image from Valigi and Mauro (2020).



The Inference Process

Project 2: Retraining or model(s) on a given cadence



Data Warehouse

