https://www.youtube.com/watch?v=T4y25jc5NyM
Creme library in Python

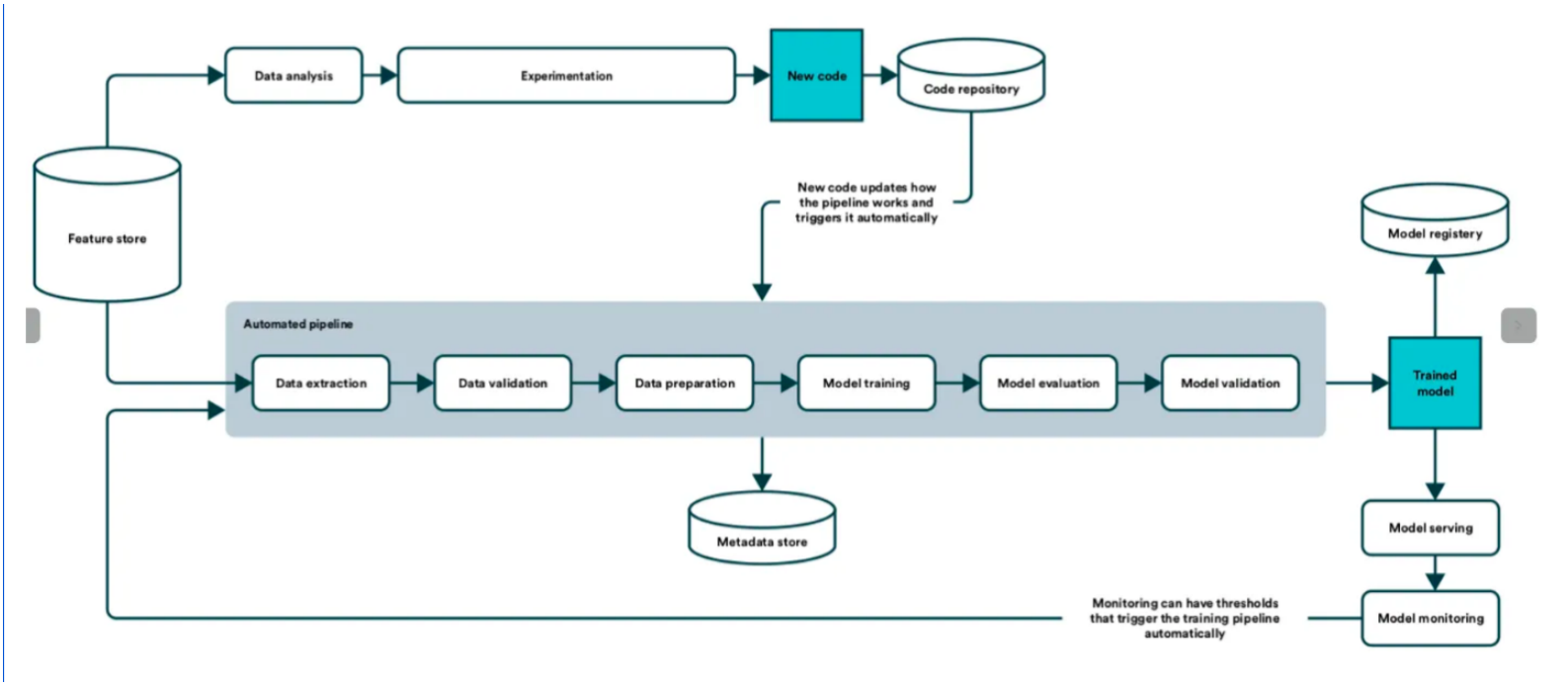
Retraining
-Incremental (online) and continuous model training

adds new data one by one

https://medium.com/@nagasanjayvijayan/continuous-training-of-ml-models-7d8acaf44dda
data drift
concept drift

1. Incremental training -training the model with new data as the data comes in (over the existing model)
2. Batch training -training the model once a significant amount of new data is available (over the existing model)
3. Retraining -retraining the entire model from scratch once a significant amount of data is available.

Automate with MLO's pipelines!!



- continuous integration
- continuous delivery
- continuous training
- continuous monitoring

Continuous Training:

1. **Data Extraction** — Extracting only the data that is needed from the data we get from the source.
2. **Data Validation —** Validating whether the data we extracted is present and is in the expected format.
3. **Data Preparation** — Processing the data to convert it into a suitable format to train the model.
4. **Model Training** — Training the Machine Learning model with the processed data.
5. **Model Evaluation** — Evaluating the metrics of the trained model.
6. **Model Validation** — Validating the new model's predictions using the old/new data and comparing it with the old model's predictions (A/B testing).

Tensorflow Extended!
Triggers for retraining (time based, when new-data arrives, changes in system, humanmade)

tools: Kubeflow, Apache Airflow, Metaflow