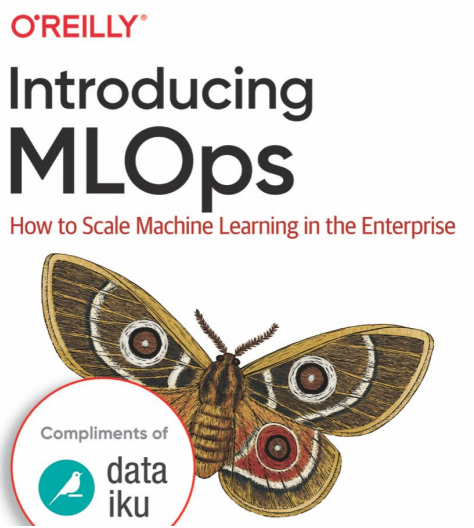
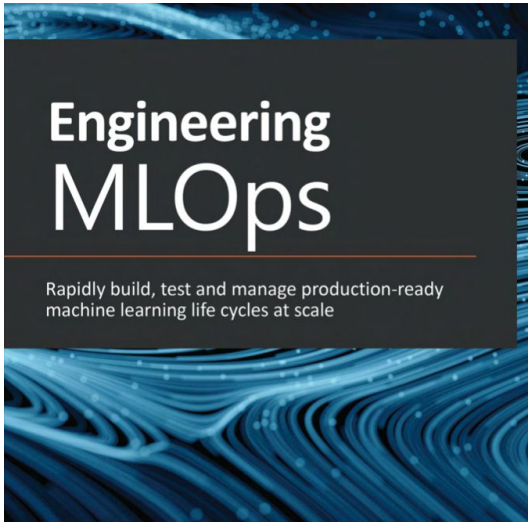
MLOps

DevOps + ML

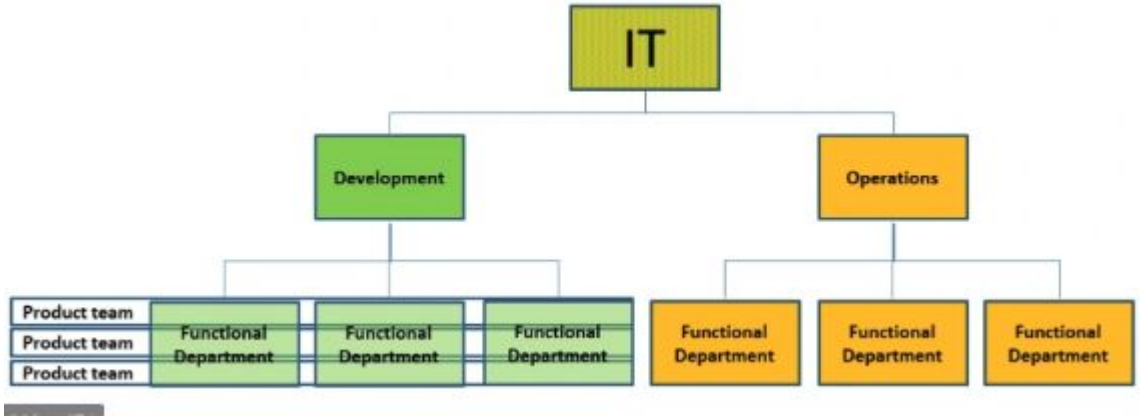
Bibliography





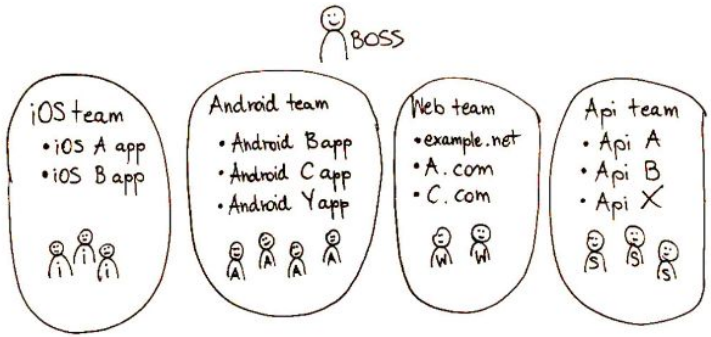
Data Products

Product Oriented IT Organization

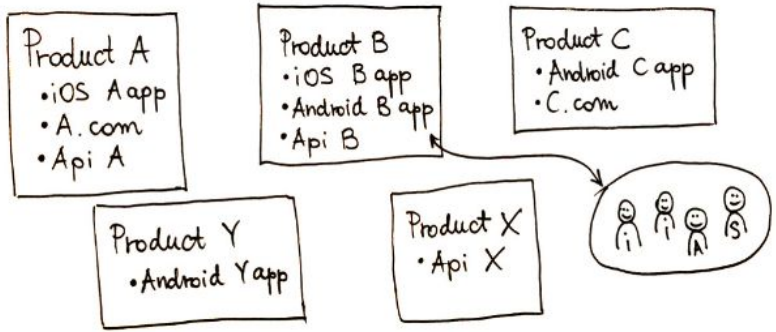


Product Oriented IT Organization

From:



To:



Source

Data Products

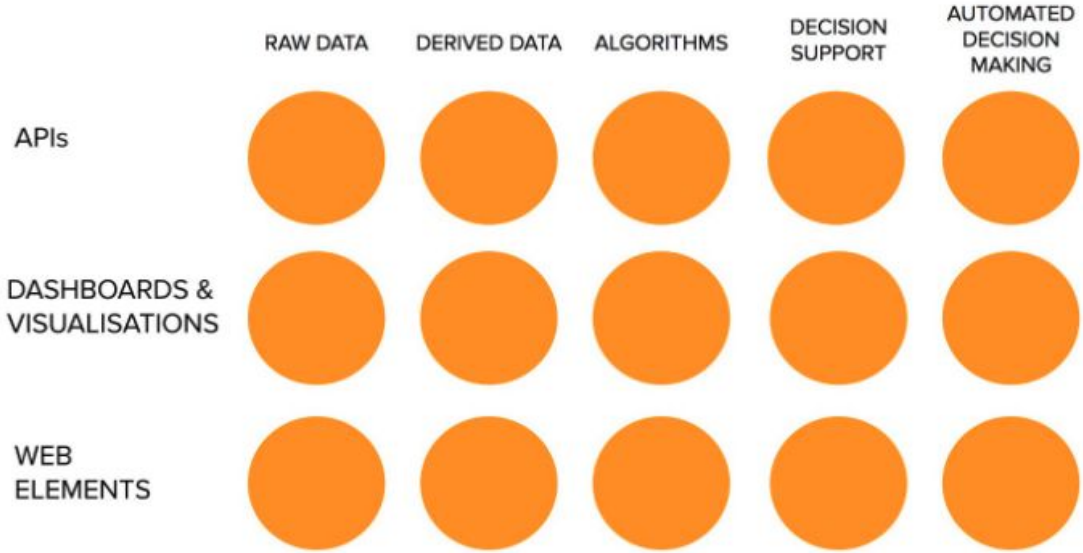
A **data product** is an application or tool that uses data to help businesses improve their decisions and processes.

We can organise these data products into **5 broad groups**: raw data, derived data, algorithms, decision support and automated decision-making.

…of data products, please

Source

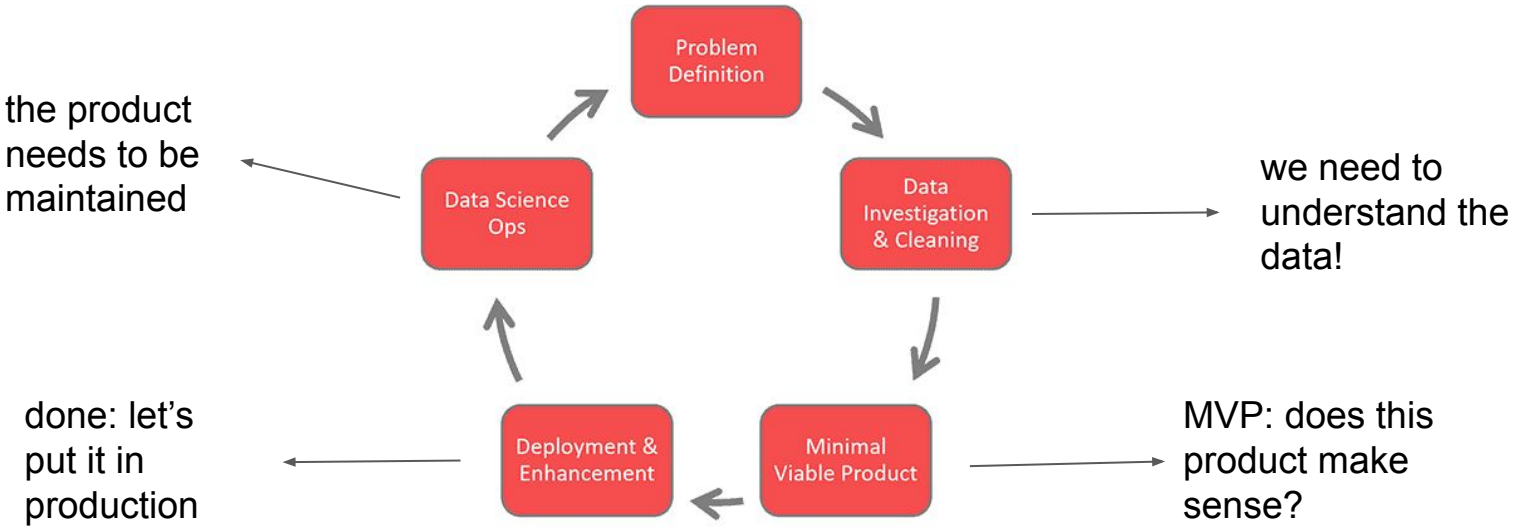
Data products: we need to understand how to consume them …



Building Data Products

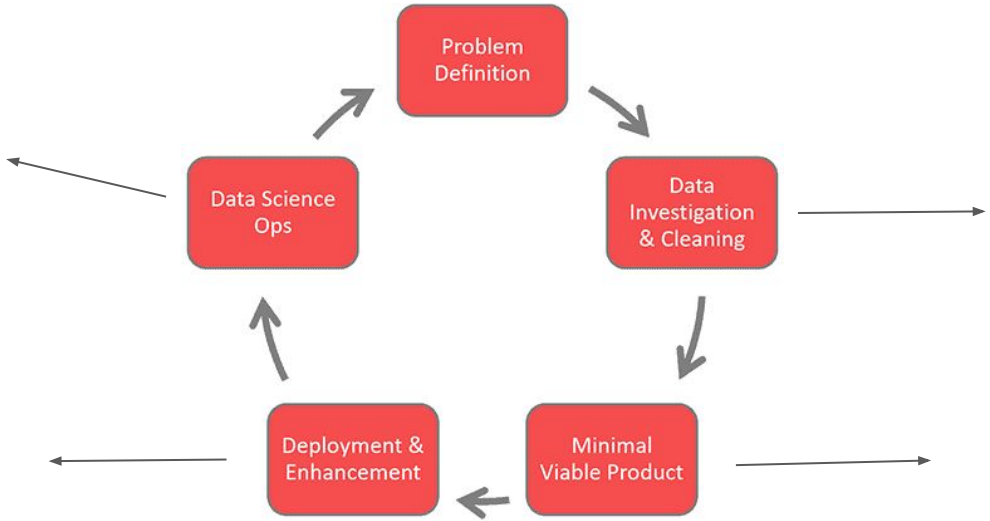
Data Science Project Lifecycle

Remember: Data Science, the way to create Data Products



Exercice: try to map it …

… to a sales dashboard in Tableau. I.e.: try to explain the actions that you should do in every phase if you have to create a sales dashboard as a data product

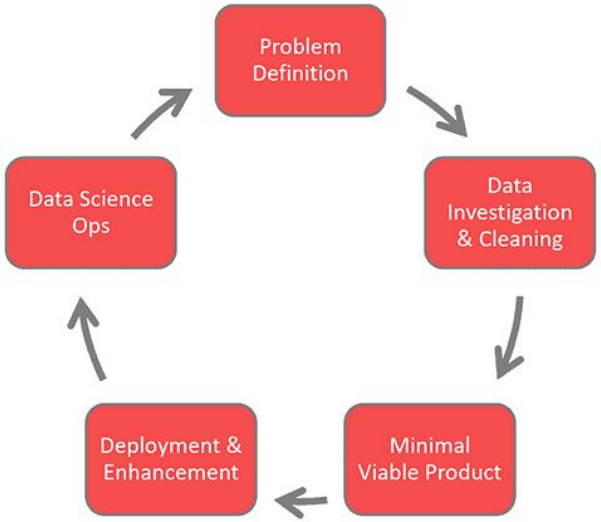


Intro to ML?

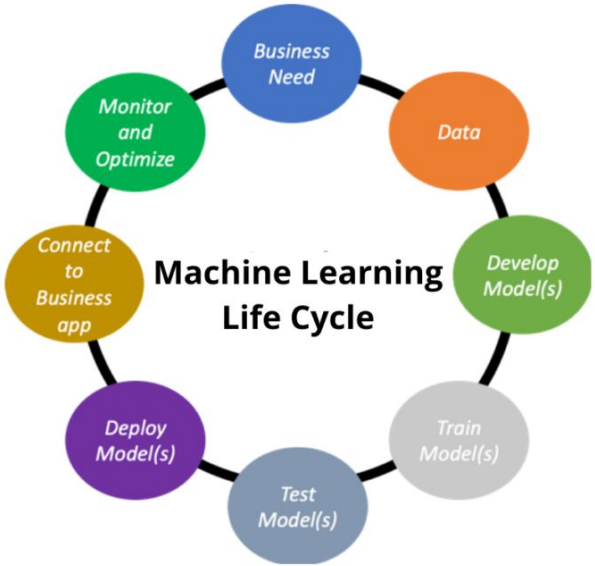
A primer in a ML Product development

The most complex data product: a Machine Learning one.

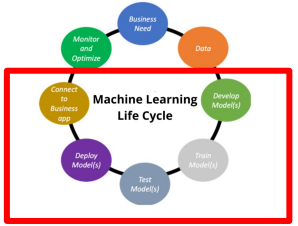
From:



To:



Zooming in…



ML Model Development Life Cycle

NoSQL Databases, HBase, HDFS, Kafka

SparkSQL, Hive, Tableau

Regularization, Grid Search, Randomized Search

Finalizing metrics and acceptance criteria → Gathering Data → Cleaning Data (ETL) → Data Analysis and Data Visualization → Training ML model → Fine Tuning ML Model → Deploying Model in Production

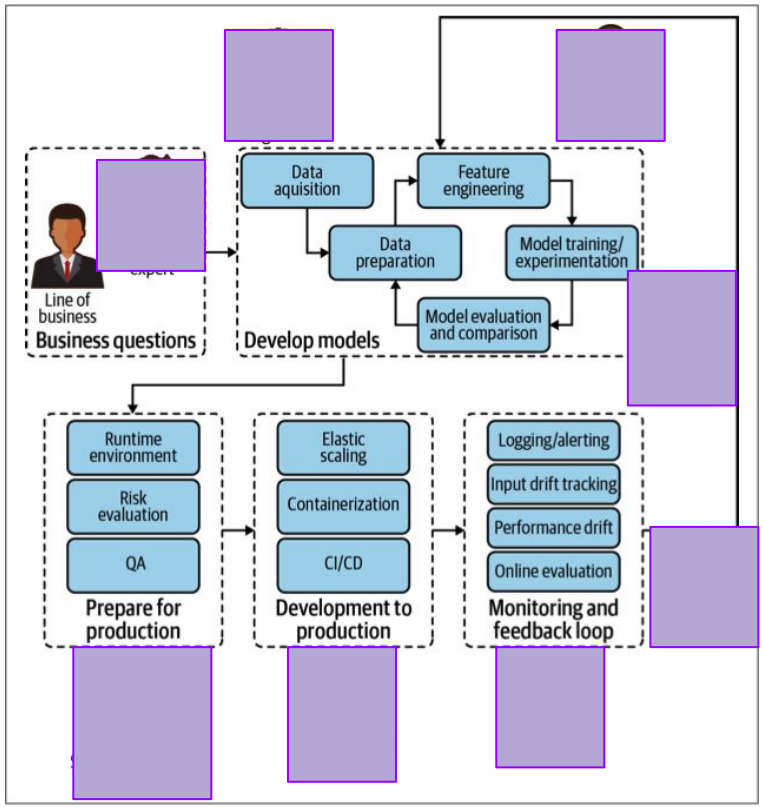
Spark RDD, MapReduce, Informatica

Spark MLLib, Scikit-learn, TensorFlow, Keras, GraphX, R

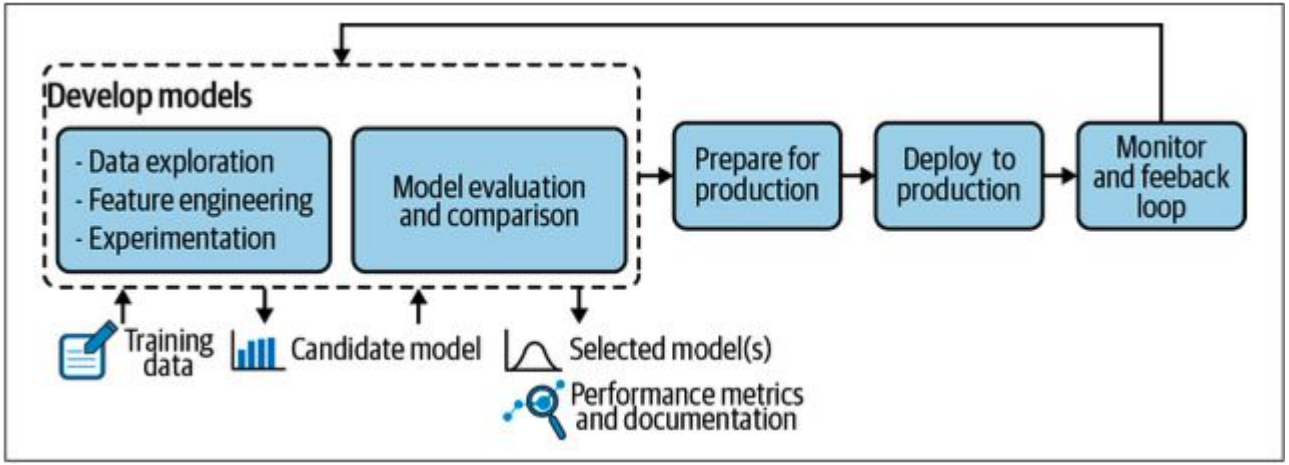
Exposing REST APIs, Building Mobile App, Devops

What is missing?

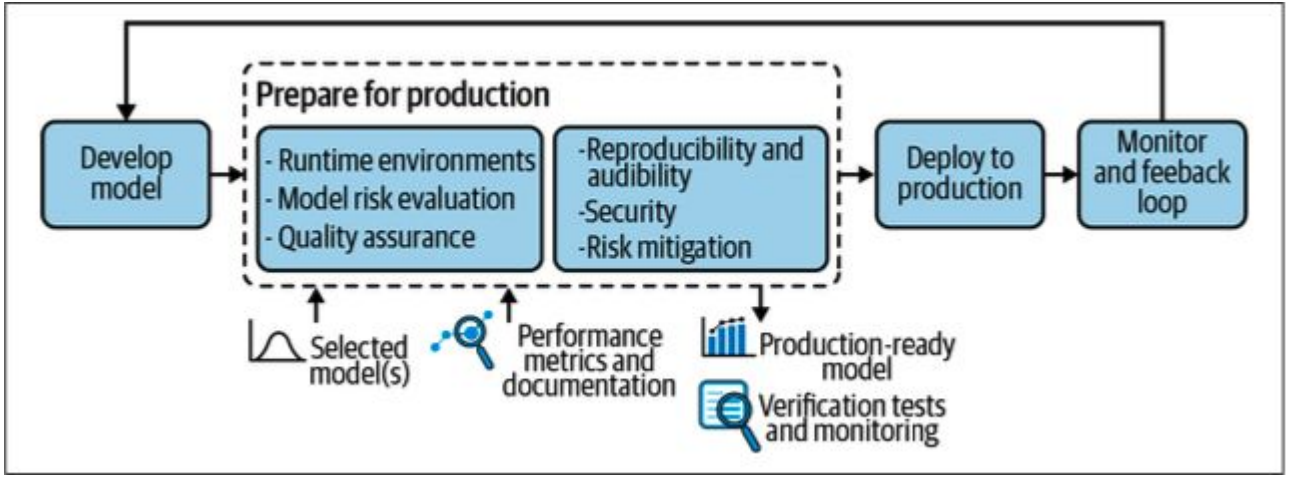
MLOps: realistic picture



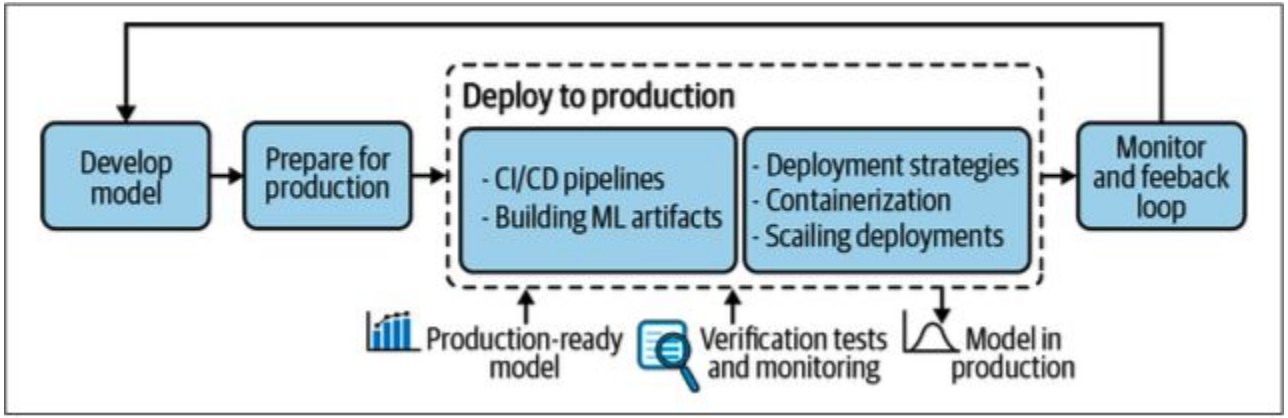
MLOps how: Step 1 - Build the ML Model



MLOps how: Step 2 - Prepare for production



MLOps how: Step 3 - Deploy to production



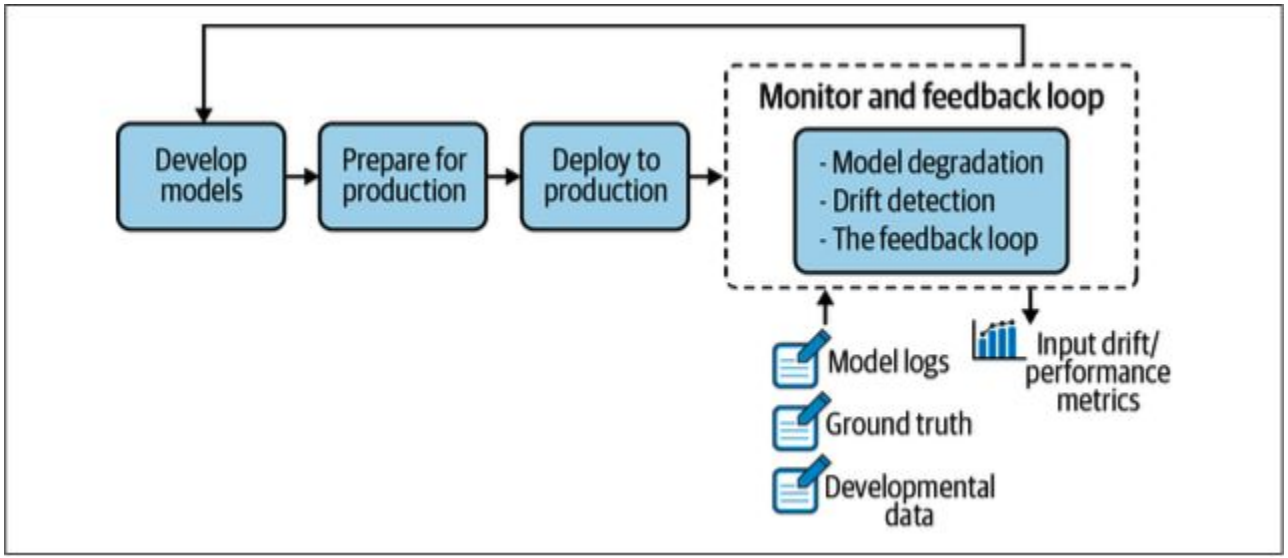
CI/CD Pipeline?

changes in:
- code
- data
- performance

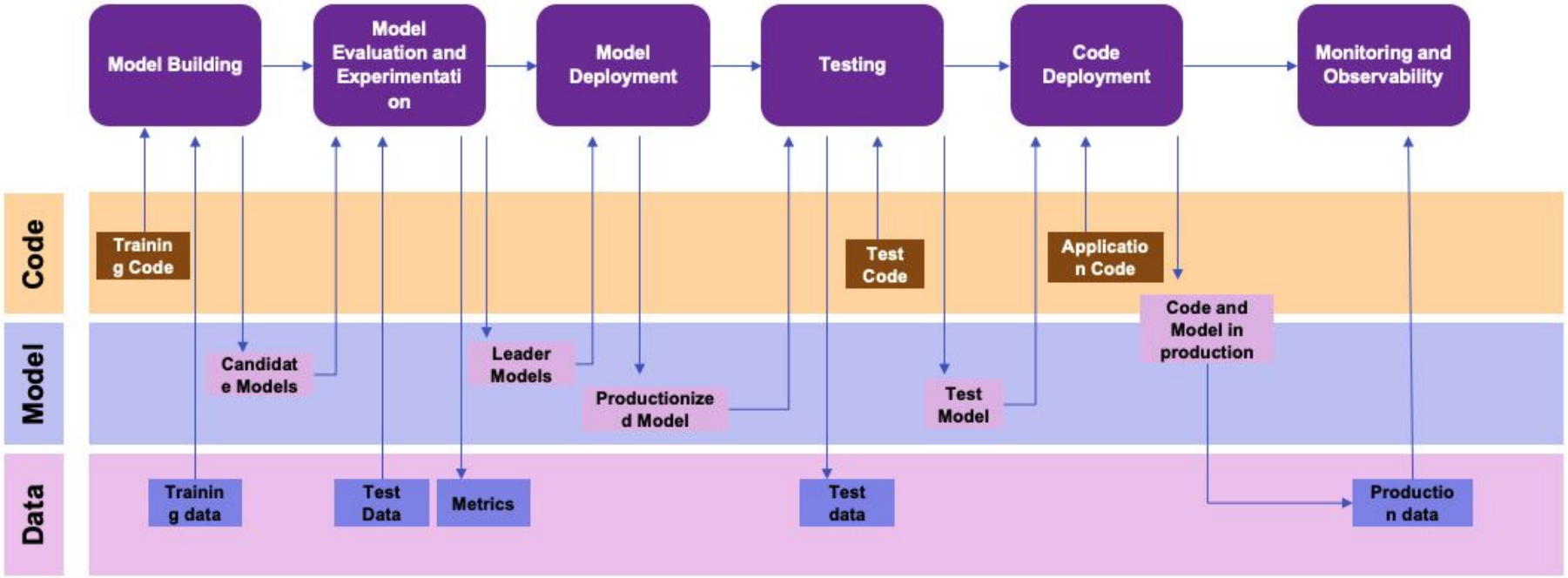
will trigger it

1. Build the model
   a. Build the model artifacts
   b. Send the artifacts to long-term storage
   c. Run basic checks (smoke tests/sanity checks)
   d. Generate fairness and explainability reports
2. Deploy to a test environment
   a. Run tests to validate ML performance, computational performance
   b. Validate manually
3. Deploy to production environment
   a. Deploy the model as canary
   b. Fully deploy the model

MLOps how: Step 4 - Monitor and feedback loop



Summary


Model Building
Model Evaluation and Experimentation
Model Deployment
Testing
Code Deployment
Monitoring and Observability
Code
Model
Data
Training Code
Test Code
Application Code
Candidate Models
Leader Models
Productionized Model
Test Model
Code and Model in production
Training data
Test Data
Metrics
Test data
Production data

Exercise

Explain the **steps** and **roles** needed to create a data product based in a ML classifier.

Problem: You want to classify customers as “good” or “bad”

Input Data: Customers and Sales tables from the Master Data Database

Performance measure: Precision

Output: The result will be a table with the id of the customer and a label

Sustain: The product has to run every night since we customer data is very dynamic.

Variability: We want to iterate on the algorithms until we find the best one, but the business wants to have something running as soon as possible. We expect data variability.

From data acquisition to performance model monitoring, explain the steps and who will be the responsible of each one.