

Premissas para estruturação de um projeto de modelo preditivo:

- respeitar uma metodologia adequada para validação de modelos
- permitir a rápida execução de todas as etapas do processo

- Ingesting and cleaning raw data sources
- Implementing governance and auditing
- Providing an environment to develop, share and collaborate on features
- Exporting training, test and validation data sets runs
- Tracking experiments, runs, hyperparameters, features, artifacts, etc.
- Testing ML models and features for performance, accuracy, and impact
- Releasing models and feature vectors as services
- Tracking lineage, model versions and performance
- Deploying batch scoring, real time serving, containers and cloud inference services
- Updating models in production as they inevitably go stale

In ML systems:

- Continuous Integration (CI) is not only about testing and validating code and components, but also testing and validating data, data schemas, and models.
- Continuous Delivery (CD) is not only about a single software package or a service, but a system (an ML training pipeline) that should automatically deploy another service (model prediction service).
- Continuous Training (CT) is a new property, unique to ML systems, that's concerned with automatically retraining candidate models for testing and serving.
- Continuous Monitoring (CM) is not only about catching errors in production systems, but also about monitoring production inference data and model performance metrics tied to business outcomes.

