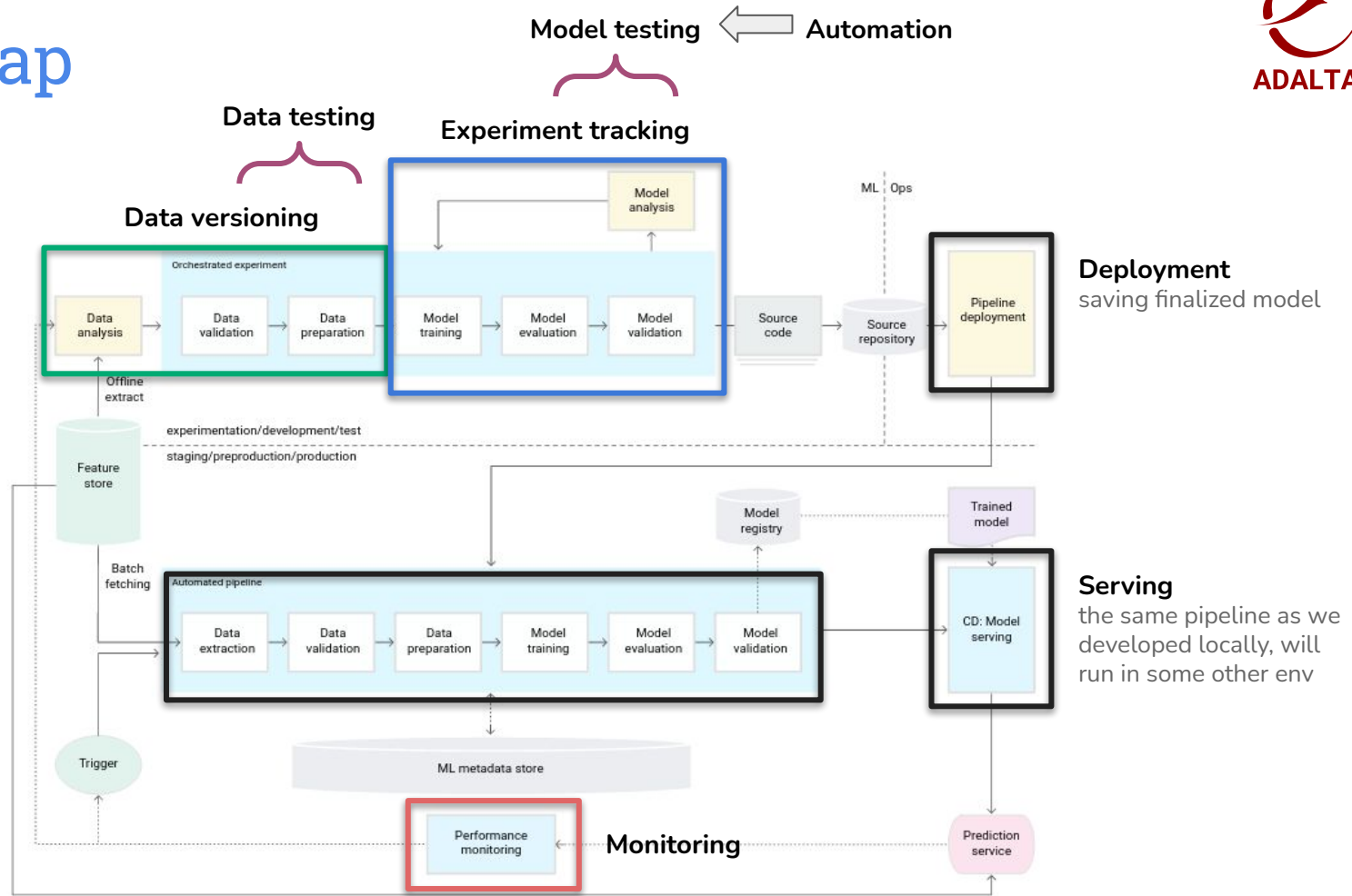


7. Monitoring

The missing chapter

Roadmap



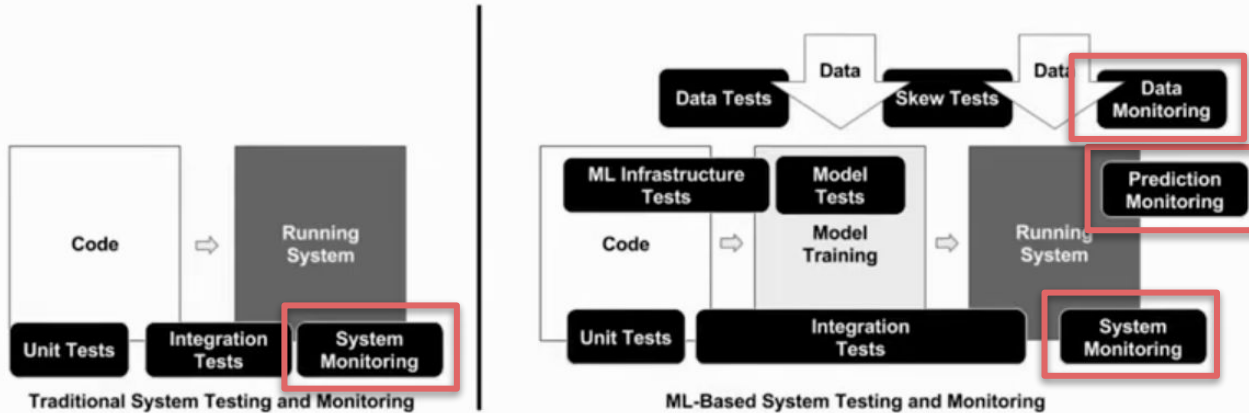
Why is it important?

- The model is in production
- We tested it thoroughly before deployment
- What can go wrong?

“Model starts degrading immediately after we put in production.”

Do you remember?

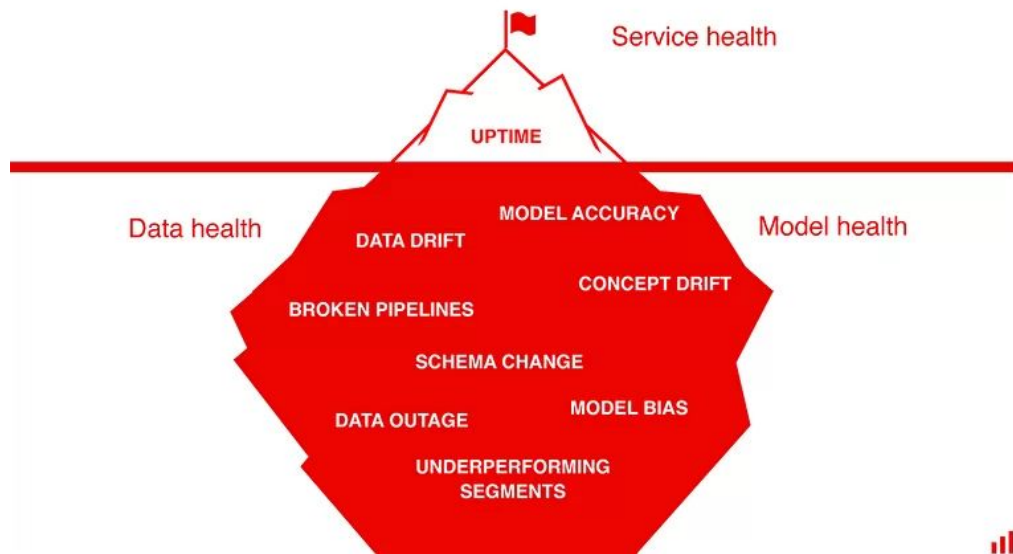
- ML in production: infra + data (pipelines) + model (pipelines)
- Things can go wrong on every step => **there's so many steps**



ML introduces two new assets into the software development lifecycle – **data** and **models**.

Extra layer of complexity

- Many failures will be silent
- You want to discover potential problems before the end-users do



System monitoring

- latency
- memory utilization
- uptime

Data monitoring - issues

How can data change with time:

- Quality issues:
 - Features (or data providers) can be discontinued
 - Change on the web page
 - Stop buying data from a specific provider
 - Content and/or type of the feature changes
 - from age to age group
 - Data get anonymised
 - GDPR
- Content issues:
 - Distribution changes
 - the data we had for development was not representative
 - changes on the market (competition, Covid, crisis, trends...)

Data quality control

Do you still remember testing?

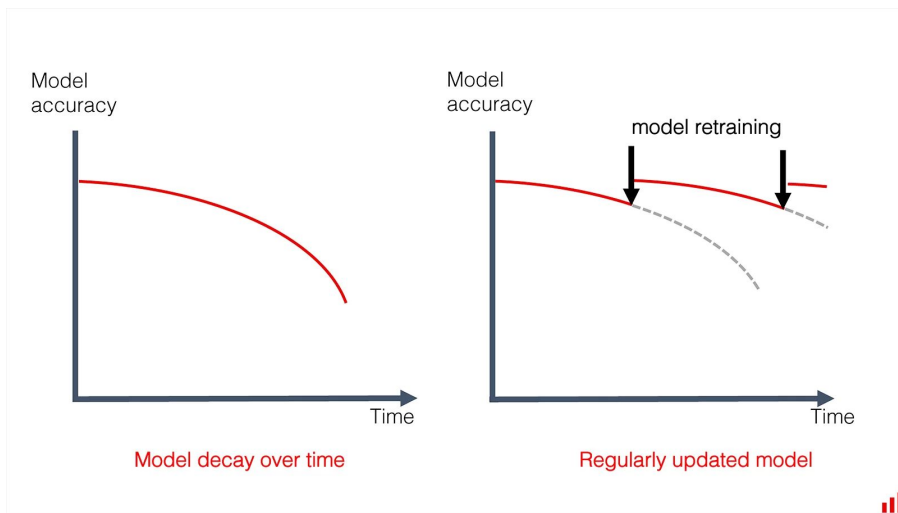
- schema
- missing values
- categories
- range
- simple statistics
- feature engineering
- ...

Data content control

- When we can think of a dataset, we can imagine two different categories of the data:
 - target variable
 - everything else
- Even though both are 'data', they don't behave the same in real time
 - availability!
- According to this properties we separate two categories of issues
 - data drift - change in the distribution of the variables
 - target drift - the relationships between the variables and the data change; business problem changes
 - **Competitors launch new products.** Consumers have more choices, and their behavior changes. As should sales forecasting models.
 - **Macroeconomic conditions evolve.** As some borrowers default on their loans, the credit risk is redefined. Scoring models need to learn it.

Model decay

- The predictions are compared to the ground truth and the performance is estimated
- All data issues lead to worse performance of the model => staleness, model decay



Data drift: examples

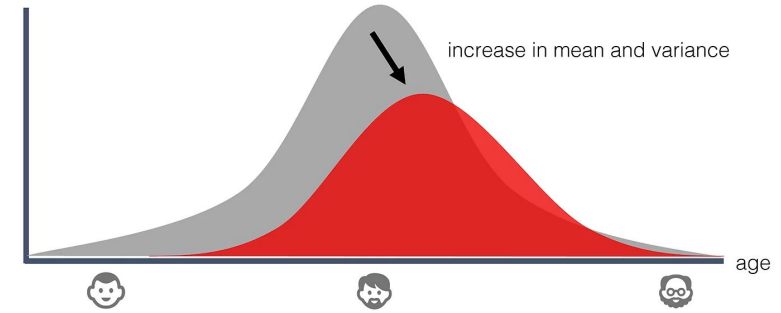
Data stream



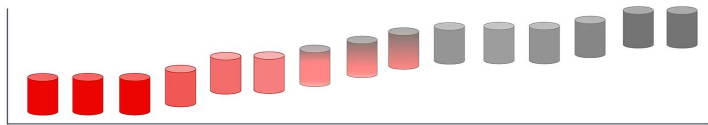
Feature distribution: source_channel



Model performance: accuracy over time

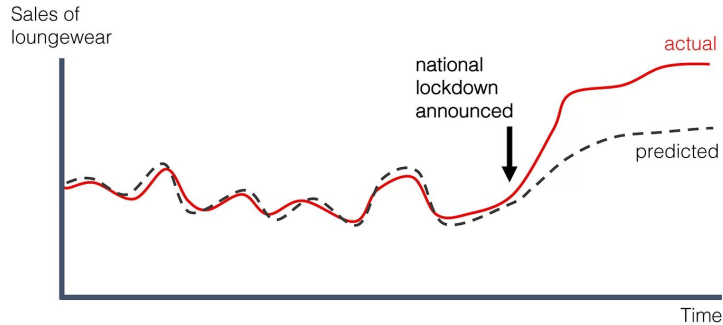
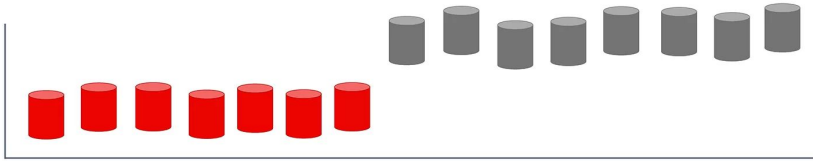


Concept drift: gradual



- **Competitors launch new products.** Consumers have more choices, and their behavior changes. As should sales forecasting models.
 - **Macroeconomic conditions evolve.** As some borrowers default on their loans, the credit risk is redefined. Scoring models need to learn it.
 - **Mechanical wear of equipment.** Under the same process parameters, the patterns are now slightly different. It affects quality prediction models in manufacturing.
-
- Consequence of evolution/adaptation
 - Normal and expected

Sudden concept drift



In a more ordinary course of events, you can experience things like:

- **Change in the interest rate by the central bank.** All financial and investment behavior is affected, and models fail to adapt to unseen patterns.
- **Technical revamp of the production line.** Predictive maintenance becomes obsolete since modified equipment has new failure modes (or lack of those).
- **Major update in the app interface.** Past data on clicks and conversion becomes irrelevant since the user journey is a new one.

What to do?

Model needs to learn new state of the world:

- retrain
- build new model

Problems:

- when is 'bad behavior' bad enough?
- is enough data available?
- what are new underlying rules?

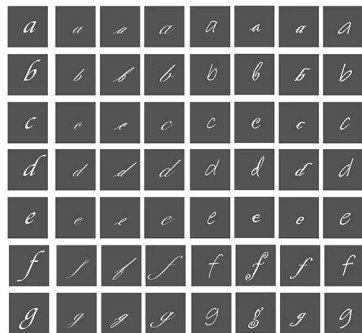
Where to go from here

Retraining:

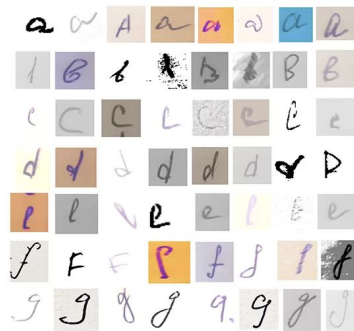
<https://evidentlyai.com/blog/retrain-or-not-retrain>

Training-serving skew (not part of the monitoring)

- Data for the training is not representative
- Test-train split



Training data



Production data

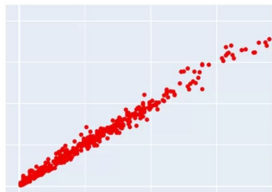


What will we do?

How to break a model in 20 days. A tutorial on production model analytics.

[Tutorial](#)

How it started



How it's going

