Production-Ready Data Science: 2017 Big Trends

DATA QUALITY

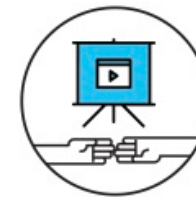**50%** agree that the biggest barrier to data deployment is data.

Data quality and pipeline development issues (and having the time available to work on data) are the number one issue.

IT CONTROLS PRODUCTION

50% do not have a specific data science production procedure.

The disconnect between data and IT team can lead to re-coding and longer design-to-production processes.

BUSINESS COLLABORATION

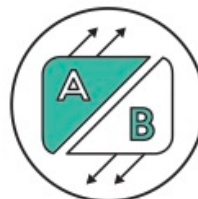**33%** of companies have close collaboration between business and data teams.

The main mode of communication on data projects remains Powerpoint or live dashboards.

KING GIT

50% of companies use classic configuration management tools, such as git.

This is representative of an IT led process with close attention to monitoring but doesn't replace a rollback strategy dedicated to data projects running in production.

A/B TESTING IS THE RULE

76% 76% report using A/B testing for model optimization.

However over half of the respondents have built a dedicated framework to perform these tests rather than look into more complex dynamic adaptive test systems.

MULTIPLE LANGUAGES AS A NORM

80% of people have a polyglot development environment.

With different team members using different technologies to fine-tune, this allows for better data products, but the skillset & technology discrepancies can complicate production processes.

TRY DATAIKU DSS

dataiku.com/dss/trynow

data iku