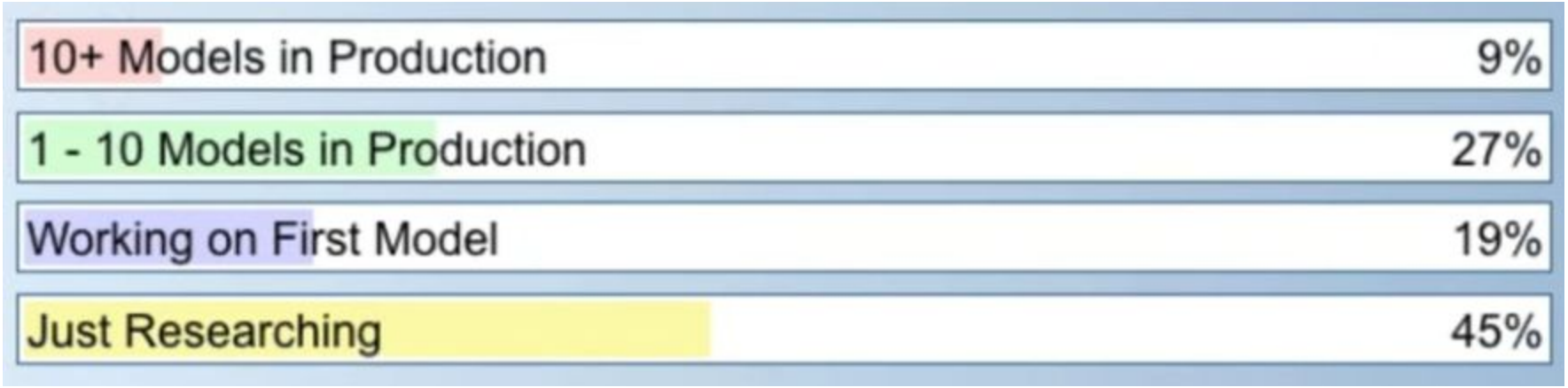
mlflow

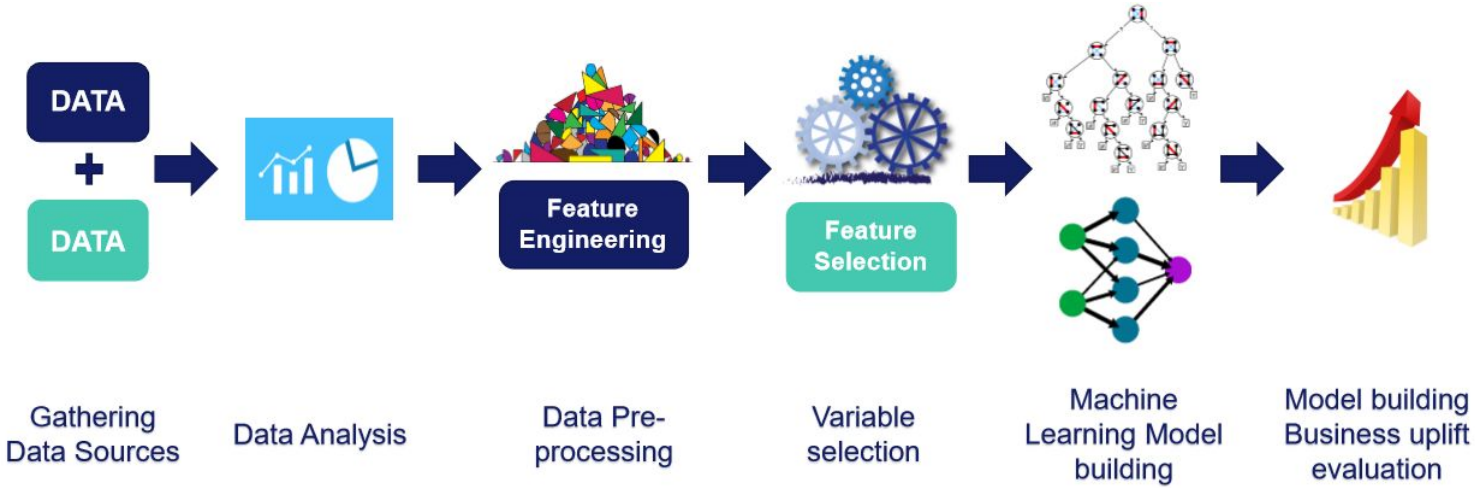
Ladle Patel
Data Scientist at Genpact

Where are you in machine learning journey ?



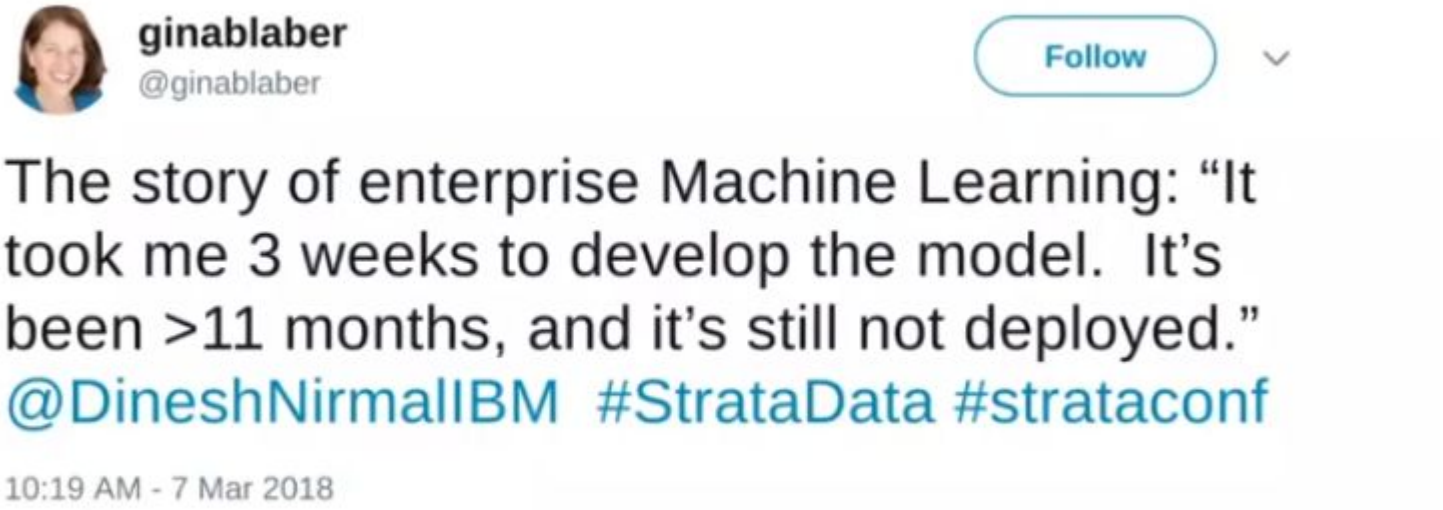
Machine Learning Pipeline

mlflow



ML In Production

ML In Production



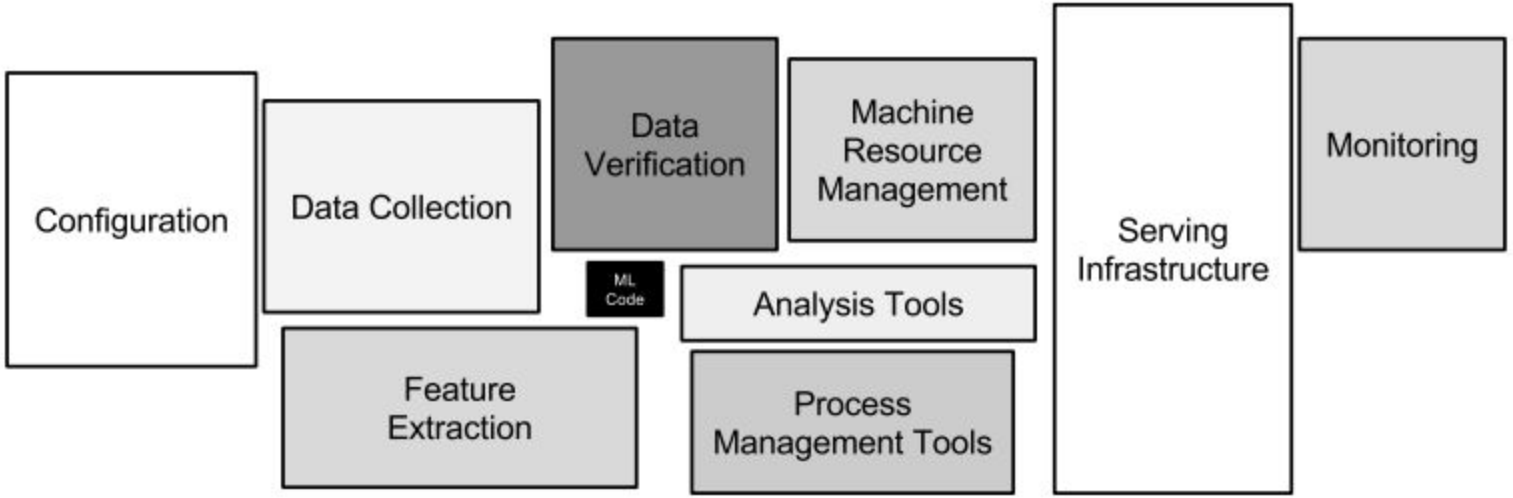
Research vs. Production Environments

mlflow

	Research	Production
Separate from customer facing software	✓	x
Reproducibility matters	Sometimes	Almost always
Scaling challenges	x	✓
Can be taken offline	✓	x
Infrastructure planning required	Sometimes	Almost always
Difficult to run experiments	x	✓

Production System

Machine learning in production requires multiple different components in order to work



Mlflow Project Motivation

mlflow

"I build 100s of models/day to lift revenue, using any library: MLlib, PyTorch, R, etc. There's no easy way to see what data went in a model from a week ago, tune it and rebuild it."

-- Chief scientist at ad tech firm

Mlflow Project Motivation

mlflow

"Our company has 100 teams using ML worldwide. We can't share work across them: when a new team tries to run some code, it often doesn't even give the same result."

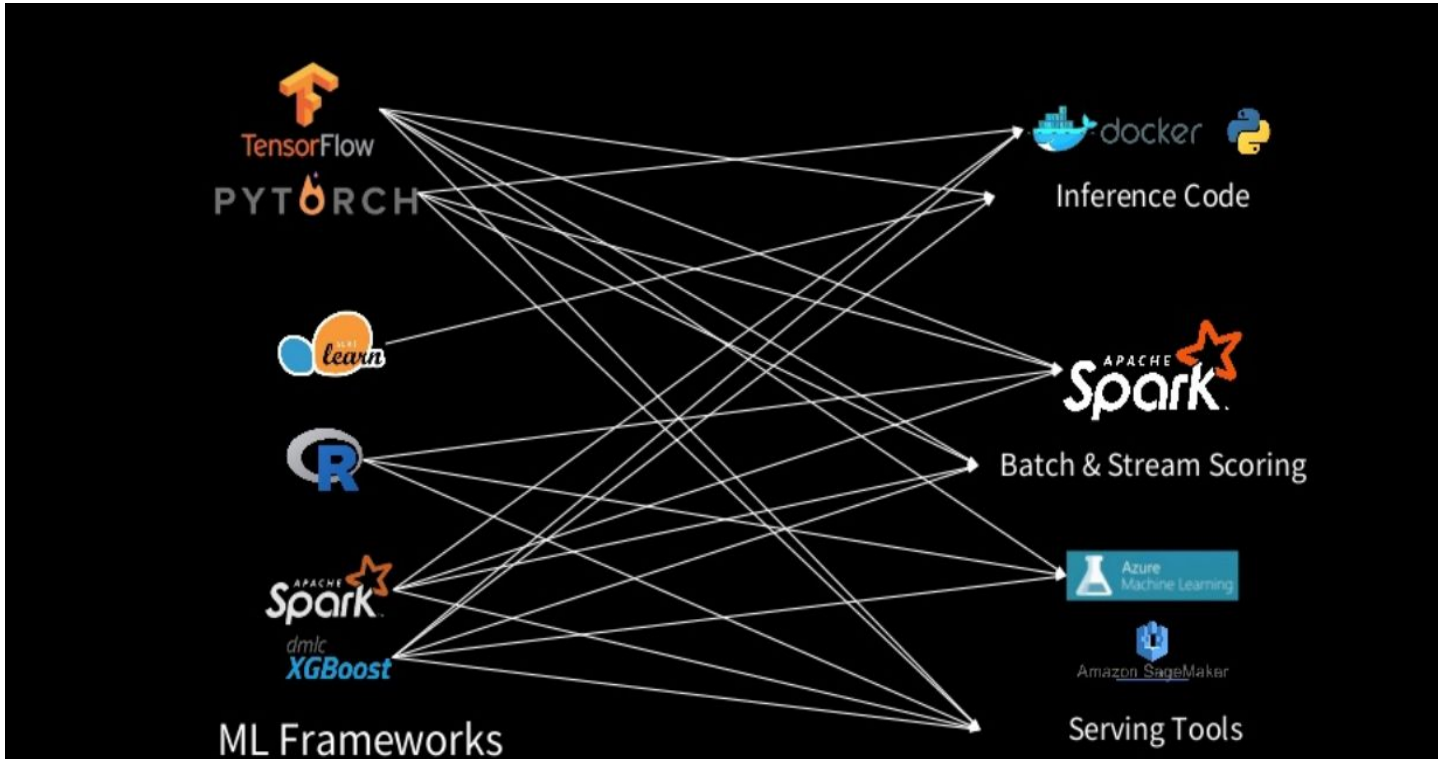
-- Large consumer electronics firm

Mlflow Project Motivation

Dev and Prod version mismatch

Mlflow Project Motivation

Diverse set of tools and environments



Mlflow Project Motivation

- **Reproducibility**: Have the ability to replicate a given ML prediction
- **Automation:** Retrain, update and deploy models as part of an automated pipeline
- **Extensibility:** Have the ability to easily add and update models.
- **Modularity:** Preprocessing/feature engineering code used in training should be organized into clear pipelines
- **Scalability:** Ability to serve model predictions to large numbers of customers (within time constraints)
- **Testing**: Test variation between model versions

What is Mlflow ?

mlflow

Open source platform to manage ML development

- Lightweight APIs & abstractions that work with any ML library
- Designed to be useful for 1 user or 1000+ person orgs
- Runs the same way anywhere (e.g. any cloud)

Key principle: "open interface" APIs that work with any existing ML library, app, deployment tool, etc

Mlflow componets

Demo

https://github.com/lp-dataninja/mlflow_learning

References

https://databricks.com/mlflow/mlflow-getting-started

https://mlflow.org/

https://storage.googleapis.com/pub-tools-public-publication-data/pdf/aad9f93b86b7addfea4c419b9100c6cdd26cacea.pdf

https://papers.nips.cc/paper/5656-hidden-technical-debt-in-machine-learning-systems.pdf

mlflow

mlflow