JAIRAM RANGANATHAN
VP Product,
KeepTruckin

#DOMINOREV

Data Science as an Organizational Capability

Building the organizational muscle for data science

Jai Ranganathan,

KeepTrucking

Who am I?

2012-2015

Product Lead - Machine Learning Platforms @ Cloudera

2015-2018

Head of Data & AI Systems (Michelangelo, Workbench, Experimentation, etc) @ Uber

Data Science Lead - Applied Machine Learning, Forecasting, Behavioral Labs, Marketing, Customer Experience @ Uber

2019-

Things we will cover

1. Team composition and scaling a data science team
2. Career growth and expectations
3. Processes for a high functioning data science organization
4. Project definition - how to make sure we are working on the right things?
5. Managing life cycles for data science projects

All the critical roles relevant for a DS org!

Data Scientist: A data analyst who lives in California

Original data science hypothesis - one person who can	**Reality**
pipeline building / ETL	Data Engineers
exploration / BI	Data Analysts (?)
modeling / ML	Data Scientists
storytelling	Product Managers
productionizing / scale	ML Engineers / Designers

Types of data scientists and designing for career growth

1. Data Scientist vs Data Analyst / BI analyst?
2. Data Science roles: Analytics, Experimentation, Modeling, Behavioral
3. Grading for growth: Not a matter of what you do but how well you do it
4. Where the rubber meets the road - Business stakeholder management and picking the right projects

Processes for a high functioning DS organization

(or how to move away from cow(boys & girls))

Development process	JIRA, Git, Version management
Data artifacts	Data versioning and reproducibility
Libraries and Abstractions	Traditional libraries AND libraries of learnings, features
Documentation	Documentation goes beyond code
Code reviews	Traditional diffs & project reviews
Experiment reviews	Design and interpretation

Processes for a high functioning DS organization

Strong Tooling!

Tooling

- Use of notebooks considered harmful
- Making compute free - big data in a single computer is very useful!
- Experimentation frameworks
- Pipeline management & feature stores
- Path to production - tools for building and monitoring

Working on the right things

Metric definitions: Trailing vs Leading

Structuring the intake process

- What metrics are you trying to move?
- What is the proposed plan of action?
- What does success look like?
- Reviewing requirements
- Exit criteria

Managing life-cycles for a data science projects

A good project never ends:

- **The world is constantly changing and our models need to keep up**
- **Sources of change:**
  - **Competition**
  - **User preference**
  - **Product changes**
- **Managing change requires:**
  - **Good measures of performance**
  - **Continuous experimentation**
  - **Tracking and alerting on performance**
  - **Good “retraining” strategy**