30K COMMUNITY MEMBERS | LIVE EVENT | TALKS & NETWORKING

DATA SCIENCE FESTIVAL *Live*

27 NOV

CODE NODE
LONDON, UK

POWERED BY DATA IDOLS.

WWW.DATASCIENCEFESTIVAL.COM

#DSFbacktolifebacktoreality

Bridging the supply-demand gap in data science

David Asboth & Shaun McGirr

Welcome

Team behind Half Stack Data Science podcast

Used to deliver data science together in the automotive sector, now do other stuff

Here today thanks to Data Science Festival!

The plan

1. Supply of data scientists
2. Demand for data science
3. What next?
4. Q&A

What is “half stack” data science?

In the half-stack world:

- Data often collected by accident
- Colleagues usually not techy
- Questions are ambiguous
- Success criteria undefined
- Interpretability matters

Takeaways:

- Data scientists need to be pragmatic problem solvers, not PhD statisticians
- This is the majority of data science (despite what you may read online)

Assuming the following is true...

1. We are churning out more aspiring/junior data scientists than ever

2. Companies are still looking for data scientists (perhaps more than ever)

… why do we keep hearing this?

1. Candidates are struggling to find the right data science job

2. Companies are struggling to find the right data scientists

People on the internet agree!

Tim Hopper
@tdhopper

I have a hard time reconciling the ostensible shortage of talent in data science/machine learning (along with the reported great employee turnover of 2021) with the stories I hear from friends of being rejected from companies after completing arduous interview loops.

12:52 AM · Nov 10, 2021 · Twitter for iPhone

25 Retweets **5** Quote Tweets **250** Likes

Part 1: problems with supply

Typical DS curriculum

Tools
EDA
Datavis
Stats
Machine learning

Typical DS curriculum

Tools
EDA
Datavis
Stats
Machine learning

Typical DS job

Data acquisition
Data contextualisation
Data “cleaning”
Ad hoc analysis
Stakeholder management
Actual analysis
Machine learning

A short aside on “data cleaning”

According to Randy Au: Data Cleaning IS Analysis, Not Grunt Work

- “80% is cleaning data” doesn’t mean 80% of time fixing date formats
- Understanding what’s behind the data takes a long time
  (requires an **analyst** skillset)
- It will be your most valuable contribution as a data scientist

Why this matters

Students on a course expect:

- Lots of “cool” machine learning
- To leave with the right skillset and get hired/deliver value

Can educators alone fix this imbalance?

Or can companies hiring data scientists help?

Part 2: problems with demand

Typical DS job advert

Tools
EDA
Datavis
Stats
Machine learning

Typical DS job advert

Tools
EDA
Datavis
Stats
Machine learning

Typical DS job

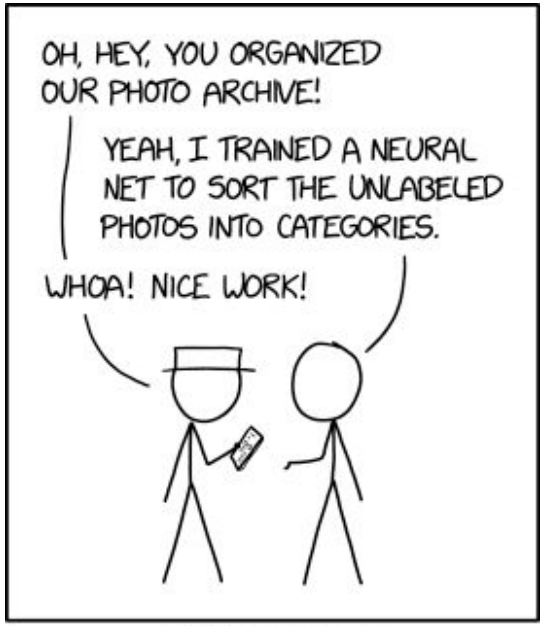
Data acquisition
Data contextualisation
Data “cleaning”
Ad hoc analysis
Stakeholder management
Actual analysis
Machine learning

Lowest hanging fruit has already fallen off the tree

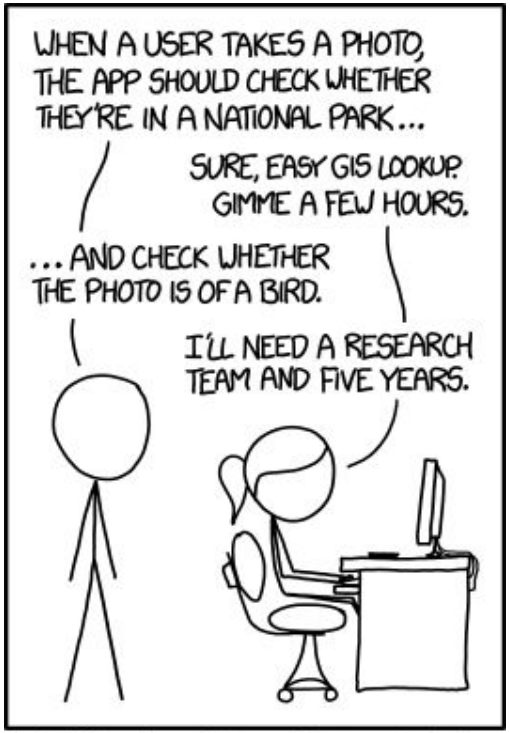
Competing with parasitic processes, not other apple pickers

Equilibrium of low expectations

xkcd unfortunately realistic



ENGINEERING TIP:
WHEN YOU DO A TASK BY HAND, YOU CAN TECHNICALLY SAY YOU TRAINED A NEURAL NET TO DO IT.



IN CS, IT CAN BE HARD TO EXPLAIN THE DIFFERENCE BETWEEN THE EASY AND THE VIRTUALLY IMPOSSIBLE.

Why this matters

Your stakeholders look out at Data Science and see:

- High failure rates
- High salaries
- High turnover

In their shoes, what would you do?

Part 3: what next?

What are these “other skills” you need to succeed?

According to Prof Roger Peng's *The Tentpoles of Data Science*

"Data Science is

1. the application of **design thinking** to data problems;
2. the creation and management of **workflows** for transforming and processing data;
3. the negotiation of **human relationships** to identify context, allocate resources, and characterize audiences for data analysis products;
4. the application of **statistical methods** to quantify evidence; and
5. the transformation of data analytic information into coherent **narratives and stories**"

We agree all of these are needed for success, but most courses only have time to teach 2 & 4

Align your concept of time & value with the organisation

Three awkward questions to ask your data leader:

1. If our slow-burning, long-running investigation in to XYZ is getting no traction, is there simply no underlying business challenge?

2. Even if our model were in production from our perspective tomorrow, how much change would it take the organisation to act differently?

3. Who owns time in our organisation and what proportion of my/team's effort should be spent in their timescale vs R&D activity?

“You’re smart people, you’ll figure it out”

Join our journey

We're only getting started!

Podcast

halfstackdatascience.com

More events, more interactive

Sign up to hear more

Online

@davidasboth

@shaunmcgirr

@HalfStackDS

Q&A

half stack
data science