Data is not flat
Working with the data is an art

Alisa Dammer
me: alisadammer.com

May 24, 2018

Structure

What do I need and how can I achieve it?

- From an idea to a MVP
- Steps required
- Things to consider beforehand

From the idea to a MVP

- Is it my core product?
    - Not that many...
- Is it an important feature?
    - Market advantage
    - Cost advantage
    - Hype advantage
- Is it a neat feature?
    - Is that feature so important?

Steps required

- Data (Legal issues, Cost, Data Mining, Big Data)
- Infrastructure (Data Engineers, Storage, DevOps, Big Data, Cost)
- Data Magic (Data Scientists, Statistics, Machine Learning, Deep Learning, Cost)
- Insights incorporation

Things to consider beforehand

- Fuzzy tasks
- Fuzzy QA
- Fuzzy results
- Cooooooost
- Data quality

I HAVE APPROXIMATE KNOWLEDGE OF MANY THINGS

Classification problem

Is it a classification or a forecast problem?
`http://localhost:8888/notebooks/classification_example.ipynb`

Forecast problem

Not the best model to use, but as an example
http://localhost:8888/notebooks/regression_case.ipynb#

Is the problem solved?

- ▶ What is feature engineering, again?
- ▶ Does it work?
- ▶ When to consider it

Thank you for attention!

Some useful information

- "Applied Predicitive Modeling" Max Kuhn, Kjell Johson
- https://machinelearningmastery.com/
  discover-feature-engineering-how-to-engineer-features-and-how-to-get-g
- https://homes.cs.washington.edu/~pedrod/papers/cacm12.pdf
- https://elitedatascience.com/feature-engineering-best-practices

Sources for images

- http://projectfreight.net/ws/project-freight-planning-part-3/
- https://www.workfromhomechristians.com.au/wp-content/uploads/2012/08/what-can-I-do1-300x300.jpeg
- https://qph.fs.quoracdn.net/main-qimg-a2bdcb6297b0091398767e4e4c69866d