BIG DATA CONFERENCE.

NOVEMBERRRR 27-29
2018

When to use Machine Learning

Tips, Tricks and Warnings

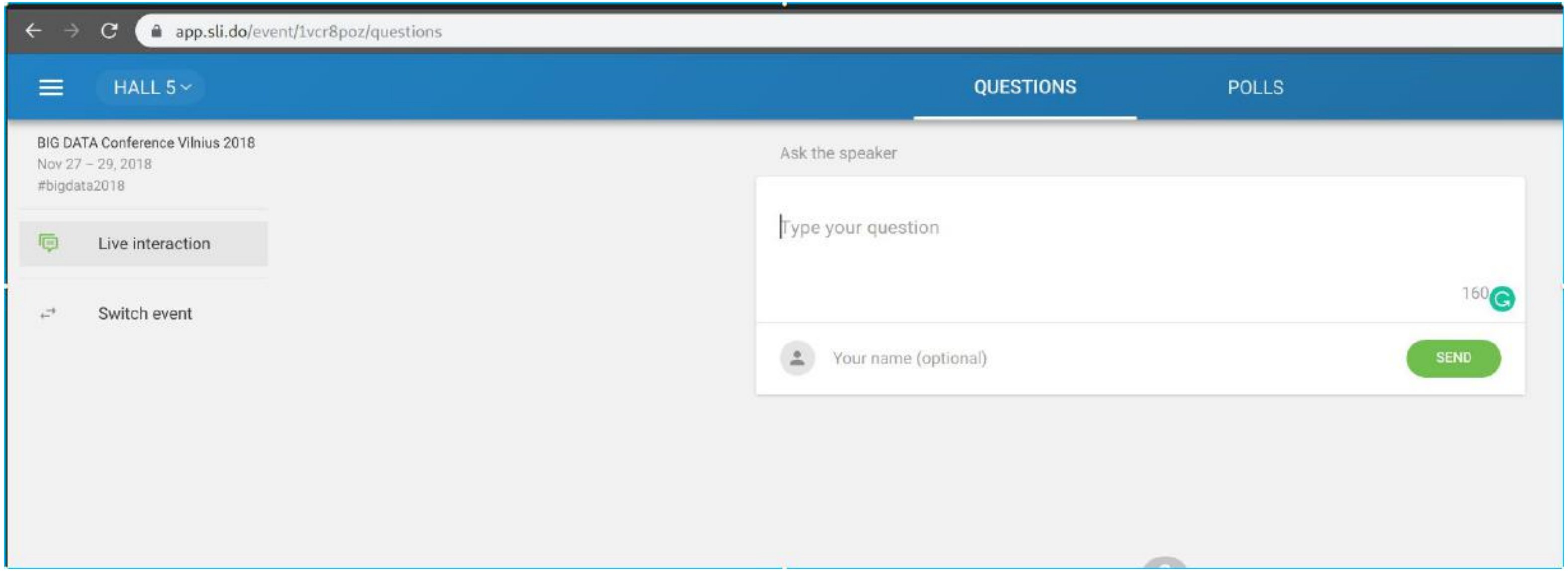
Pascal van Kooten

Jibes®

For questions: https://sli.do

Join with: bigdata2018

Room: Hall 5



BIG DATA Conference Vilnius 2018
Nov 27 – 29, 2018
#bigdata2018

Ask the speaker

Type your question

160

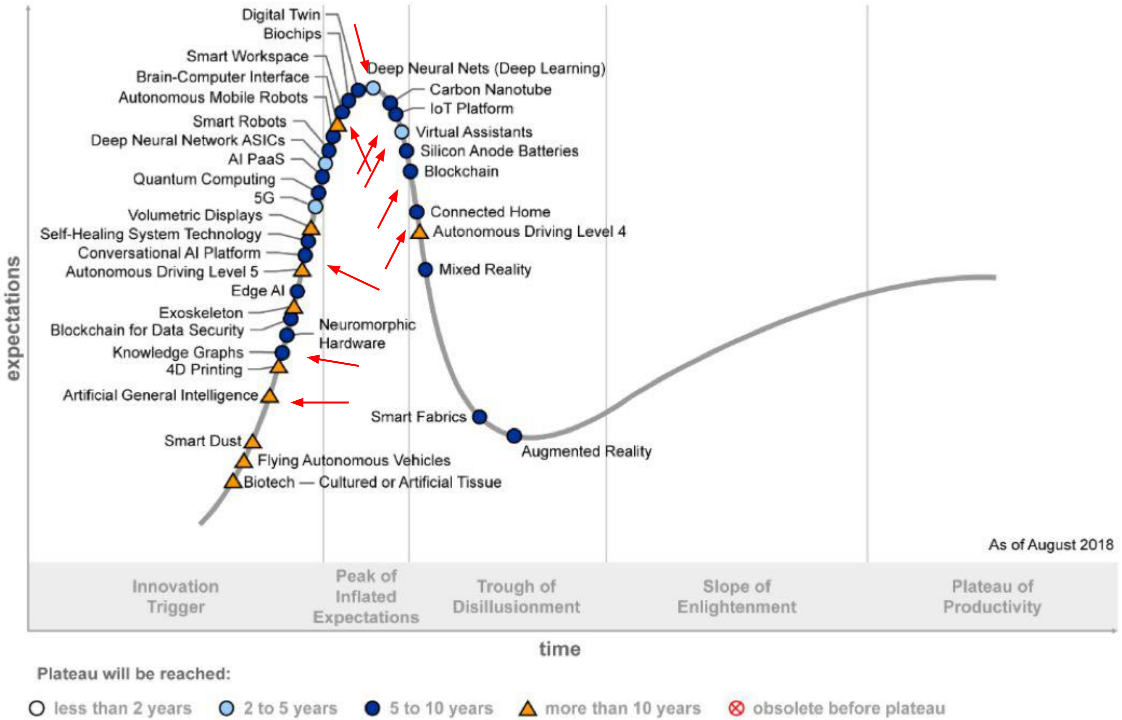
Your name (optional)

whoami

- MSc Methods & Statistics
- Intel AI Innovator
- Kaggle Master
- Lead Data Scientist @ **Jibes Data Analytics**
  - 35 data scientists
  - 4 years and 15+ different companies
  - Worked on blockchain, NLP, ML/DL, social robots
- Loves:
  - Open-source
  - Tech Innovation
  - Human & Machine Interaction

Projects

github.com/kootenpv
kootenpv.github.io



© 2018 Gartner, Inc.

Today

- What is Machine Learning *exactly*?*
- When to use Machine Learning?*
    - Example driven

* Might contain traces of code

What is machine learning *exactly*?

Learning what Machine Learning is (by data)

age	**income**
20	20000
30	30000
40	40000
50	50000
60	?????

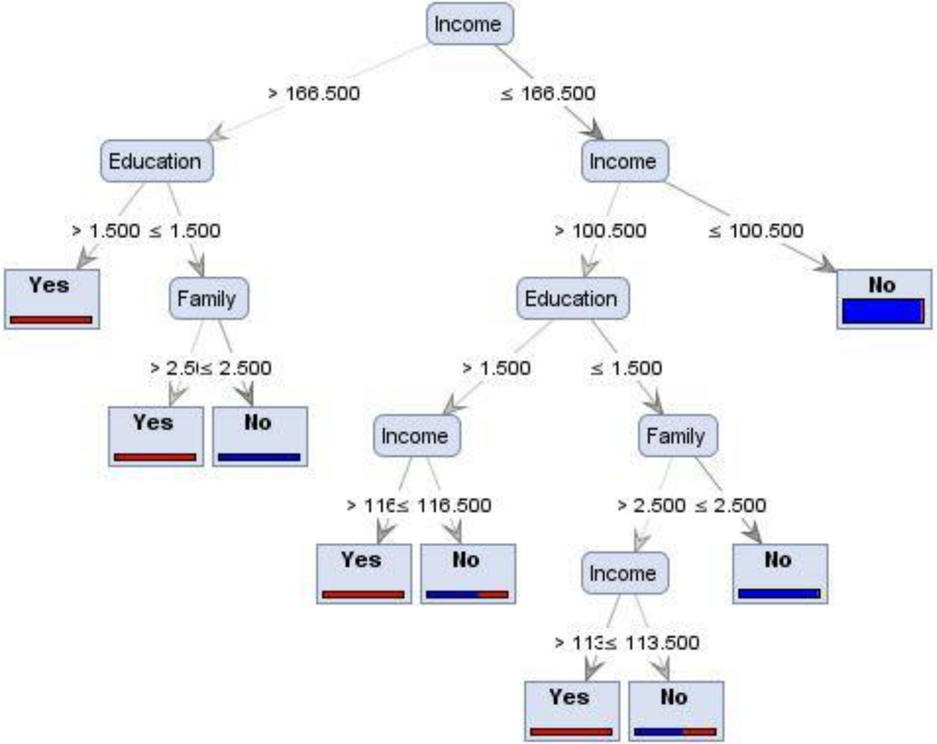
Learning what Machine Learning is (by data)

age	**income**
20	20000
30	30000
40	40000
50	50000
60	60000

Learning what Machine Learning is (by data)

age	**income**
20	**20000**
30	**30000**
40	**40000**
50	50000
60	60000

Decision Tree



Source: SimaFore

Machine Learning

- Predict whether email is spam or not:

email	**spam/not spam**
Hi John, how are you?	not spam*
Click link for FREE … !!	spam

- Rather than write a lot of if/else statements
- Learn logic based on existing input/output examples

Learning what Machine Learning is (by data)

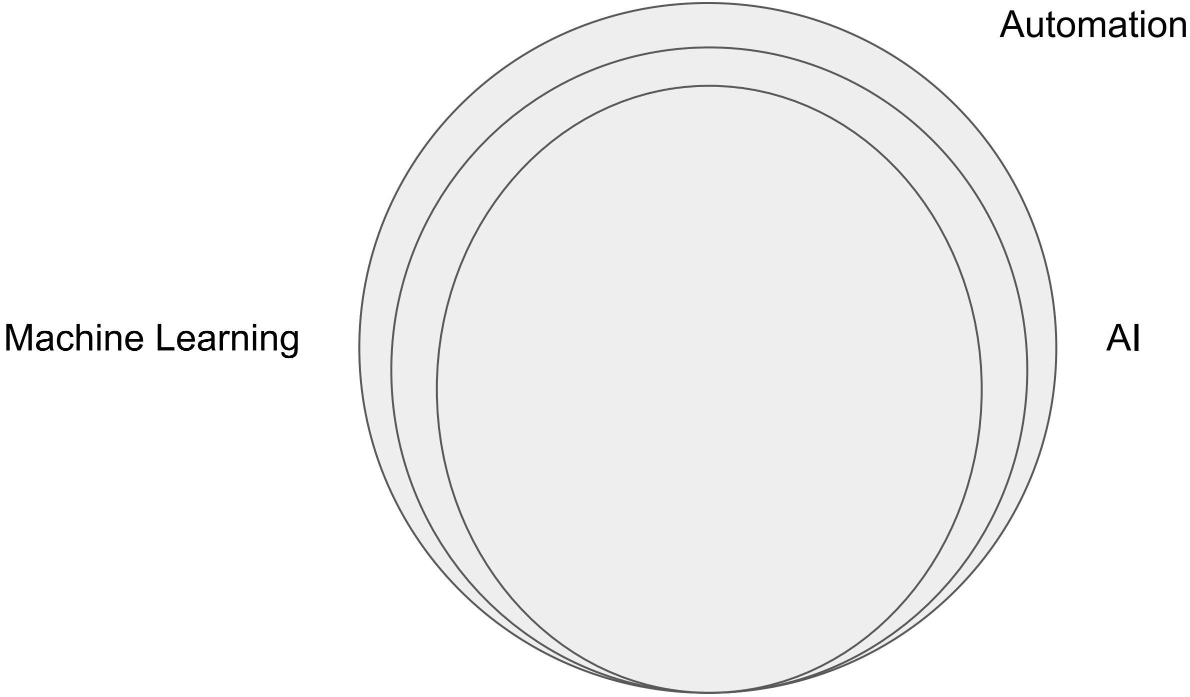
1. Find a problem (X, y)

2. Preprocess

3. Find model(s)

4. Use best model in Prod

The Significance of Machine Learning

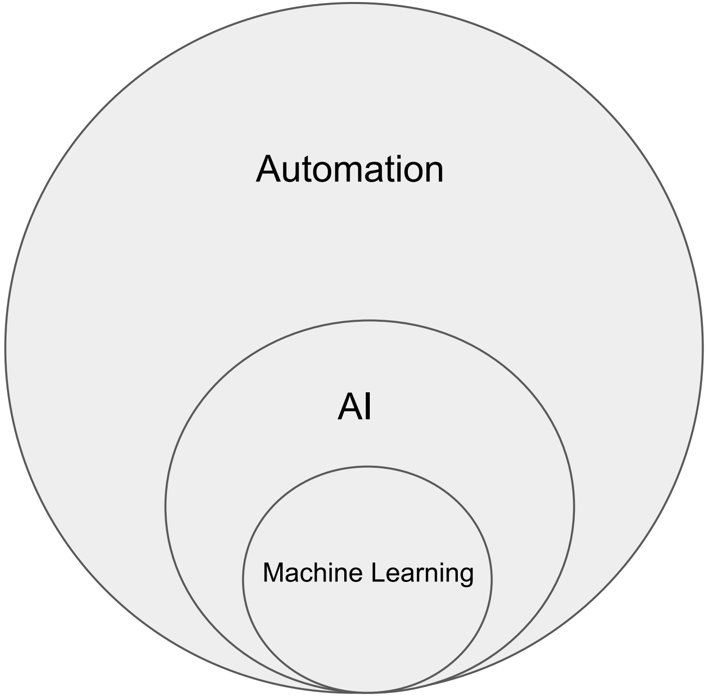


Automation

Machine Learning

AI

The Significance of Machine Learning



When to use Machine Learning

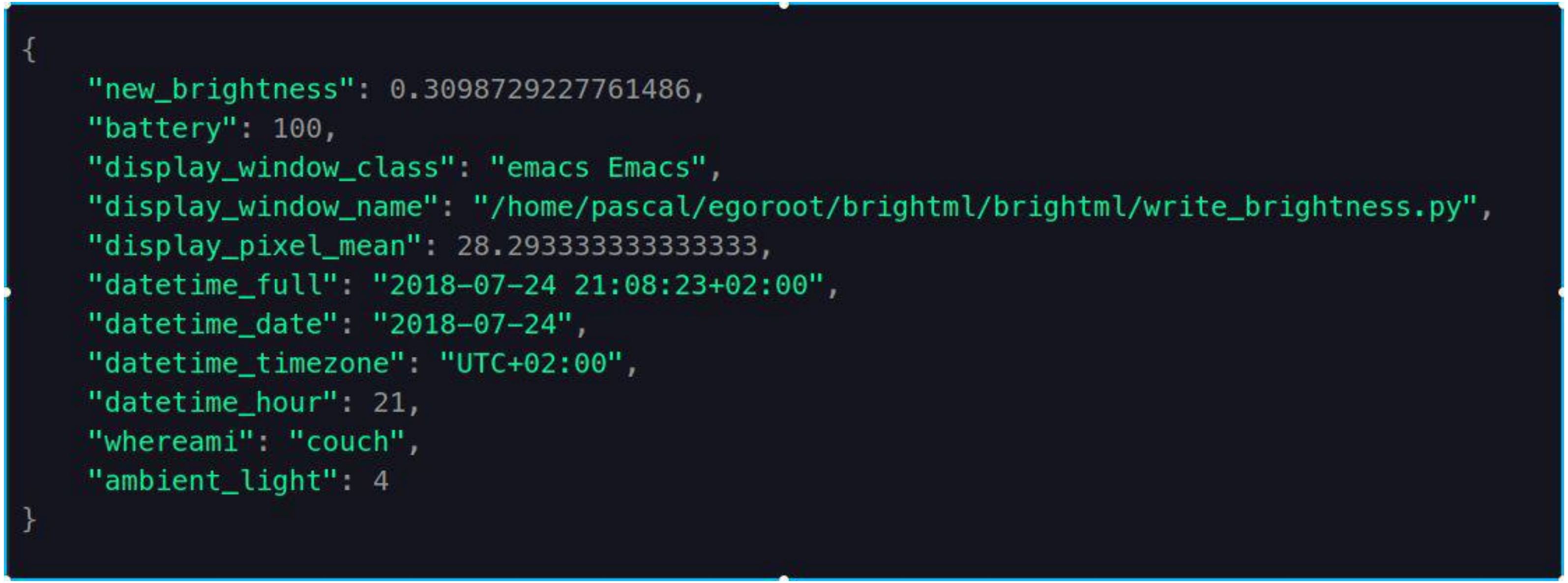
bright**ml**

Convenient Machine-Learned Automatic Brightness

DEMO

https://github.com/kootenpv/brightml

bright**ml**



```
{
    "new_brightness": 0.3098729227761486,
    "battery": 100,
    "display_window_class": "emacs Emacs",
    "display_window_name": "/home/pascal/egoroot/brightml/brightml/write_brightness.py",
    "display_pixel_mean": 28.2933333333333333,
    "datetime_full": "2018-07-24 21:08:23+02:00",
    "datetime_date": "2018-07-24",
    "datetime_timezone": "UTC+02:00",
    "datetime_hour": 21,
    "whereami": "couch",
    "ambient_light": 4
}
```

bright**ml**

- Feedback loop not noticeable
- Zero config while still personalized
- New features can easily be added

whereami

Uses wifi signal and machine learning to predict where you are

DEMO

https://github.com/kootenpv/whereami

whereami

- Pluggability is key
- Easier to learn from observation vs coding rules vs config

Computer Vision

X (image)	**y (dog)**
80x80x3px	0
80x80x3px	1

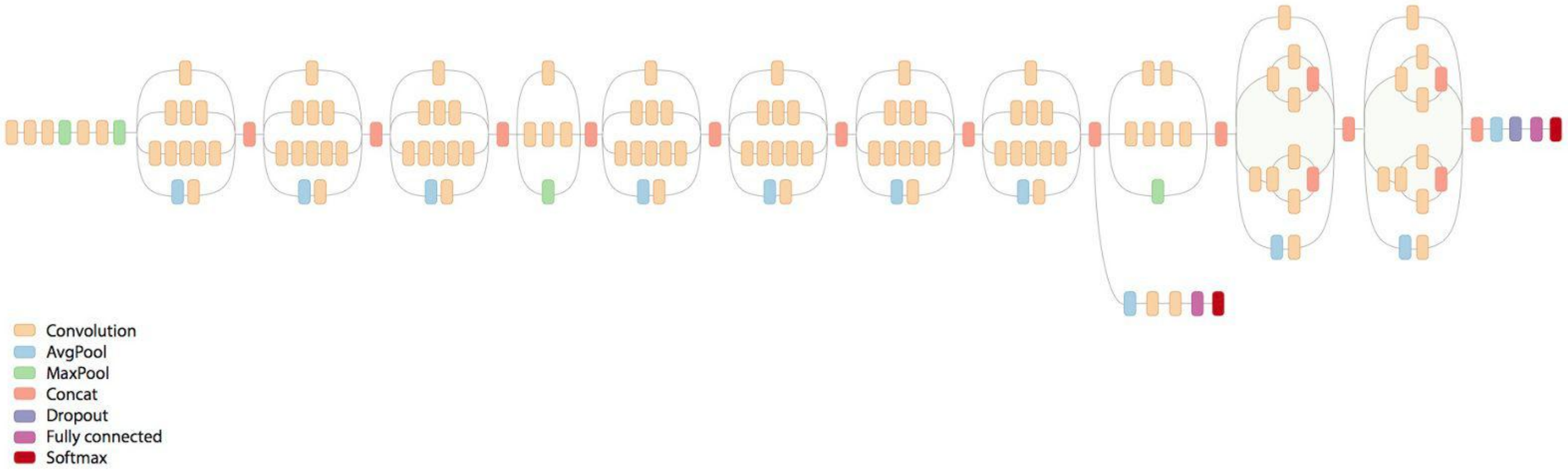
Insurance company

- Large broad insurance company
- **Investigate** what Computer Vision could do for them
- Task: predict damage $$$ from damaged car pictures

Insurance company

Insurance company

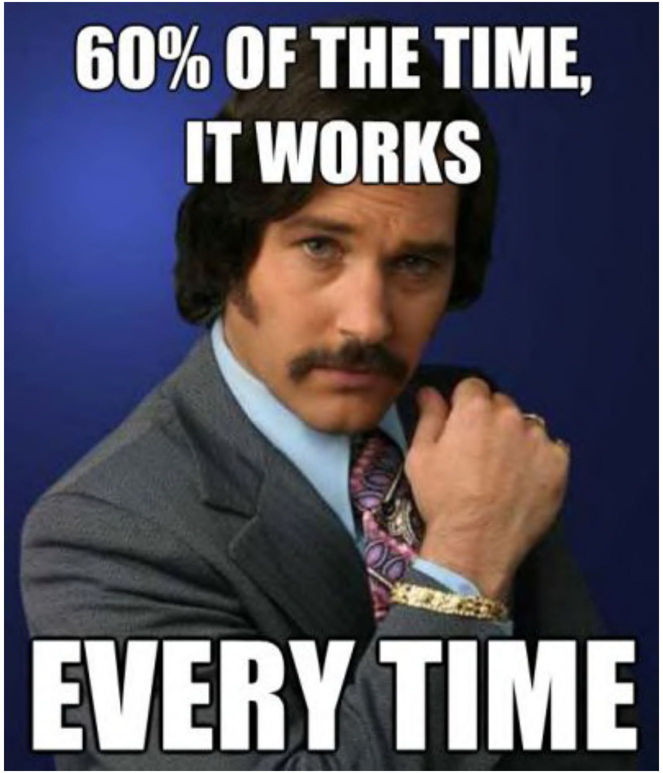
Insurance company: transfer learning



Insurance company

- Strict rules already in place
- Transfer learning can help
- But…due to complexity... data...data...data…

Insurance company



Neural Complete

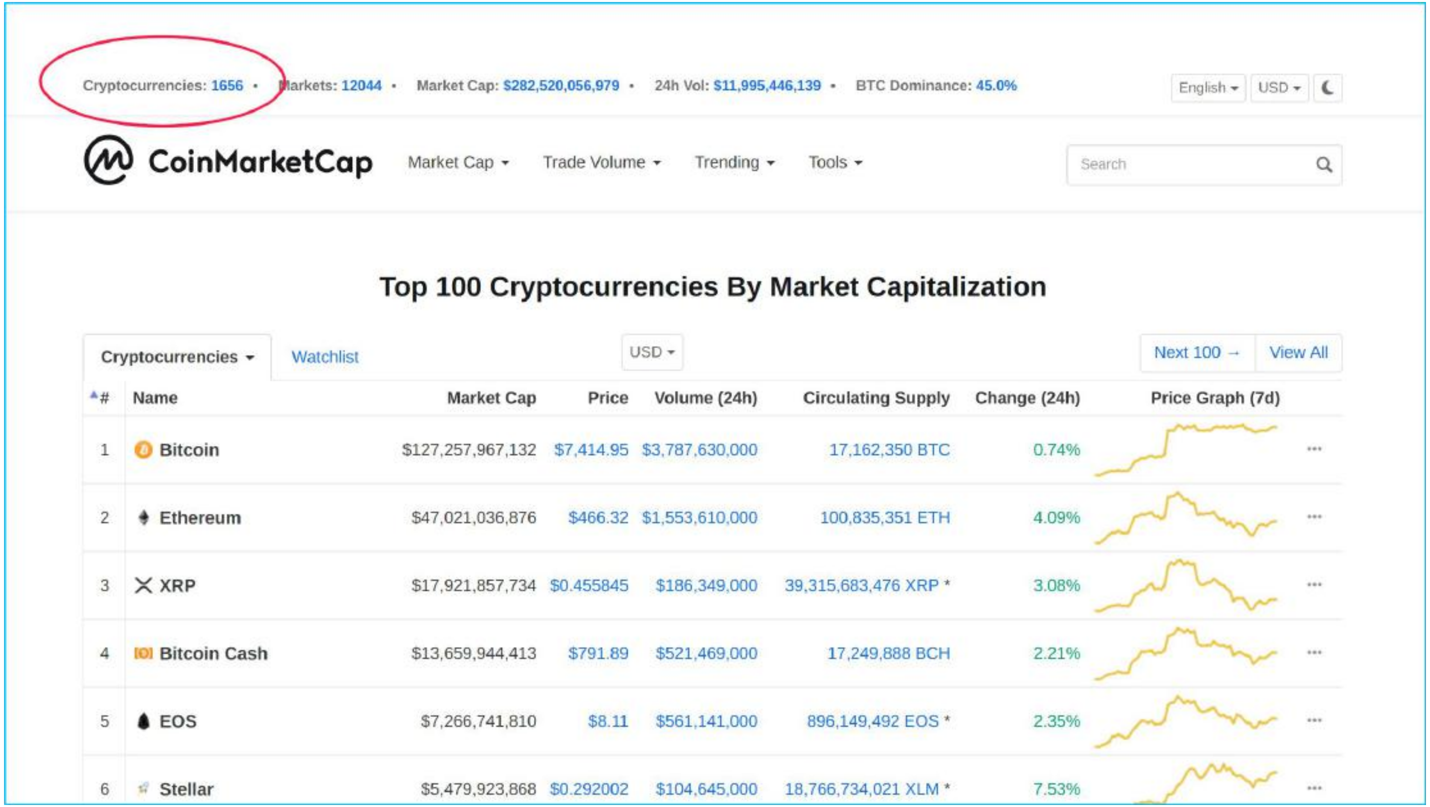
neural complete!

Choose a model: neural_token

```
# generated code follows below
```

https://github.com/kootenpv/neural_complete

Cryptocurrency Trading

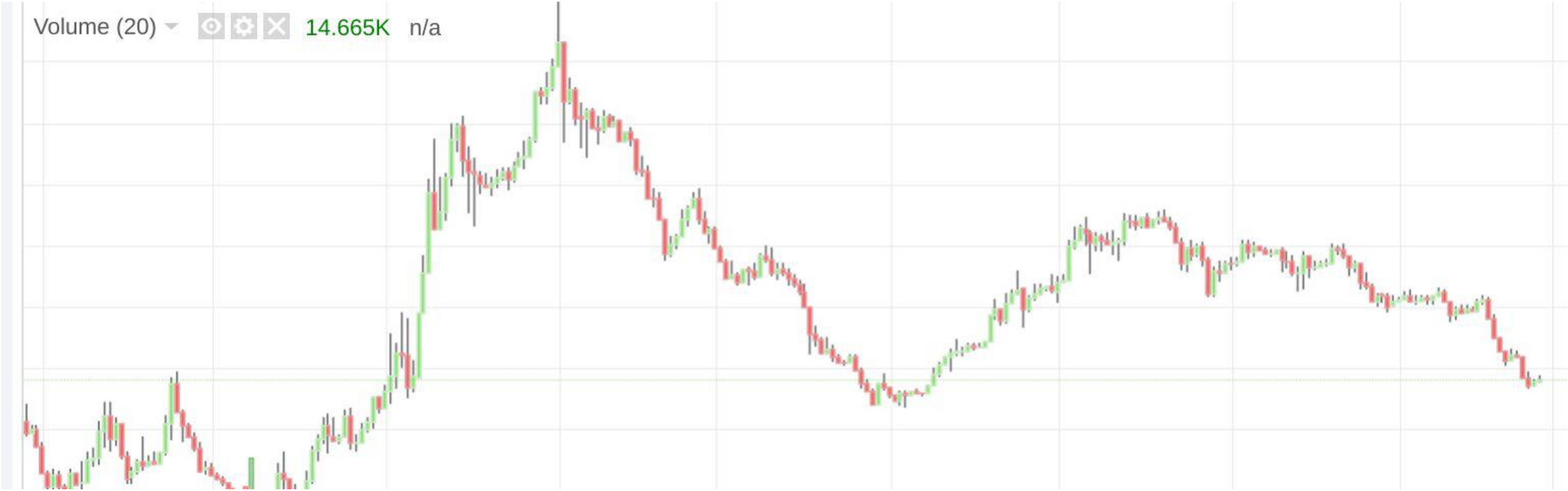


Cryptocurrencies: 1656 • Markets: 12044 • Market Cap: $282,520,056,979 • 24h Vol: $11,995,446,139 • BTC Dominance: 45.0%

Top 100 Cryptocurrencies By Market Capitalization

#	Name	Market Cap	Price	Volume (24h)	Circulating Supply	Change (24h)	Price Graph (7d)	
1	Bitcoin	$127,257,967,132	$7,414.95	$3,787,630,000	17,162,350 BTC	0.74%		···
2	Ethereum	$47,021,036,876	$466.32	$1,553,610,000	100,835,351 ETH	4.09%		···
3	XRP	$17,921,857,734	$0.455845	$186,349,000	39,315,683,476 XRP *	3.08%		···
4	Bitcoin Cash	$13,659,944,413	$791.89	$521,469,000	17,249,888 BCH	2.21%		···
5	EOS	$7,266,741,810	$8.11	$561,141,000	896,149,492 EOS *	2.35%		···
6	Stellar	$5,479,923,868	$0.292002	$104,645,000	18,766,734,021 XLM *	7.53%		···

Cryptocurrency Trading



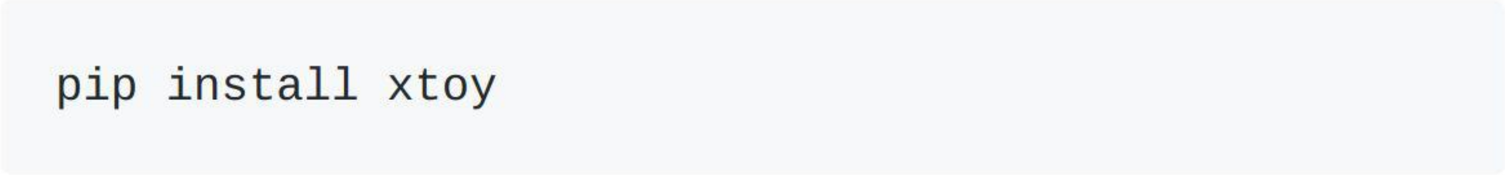
Cryptocurrency

- Don’t underestimate the work necessary next to machine learning
- Analysis vs machine learning
- Simple is better than complex 🤣

xtoy (automated machine learning)

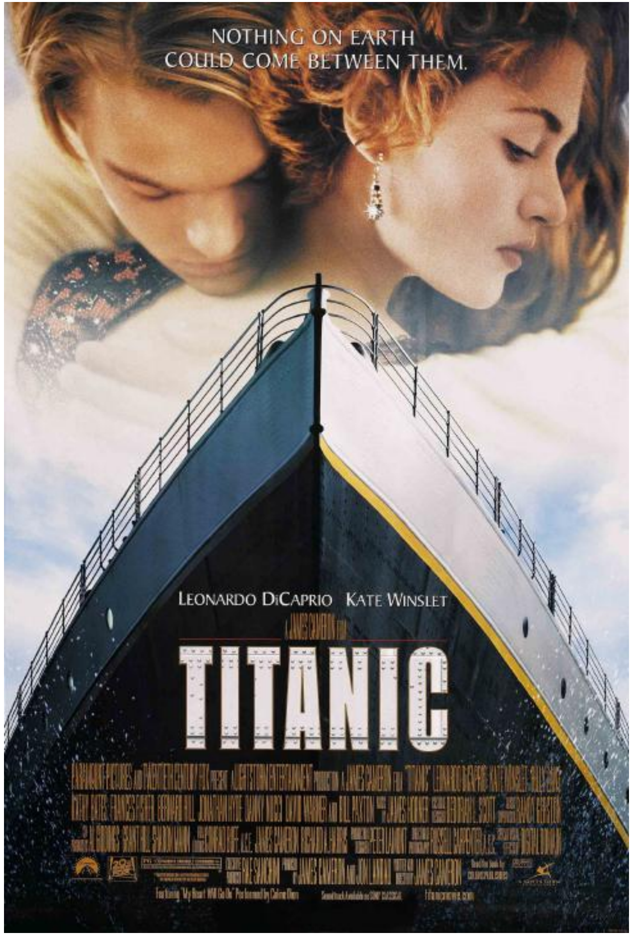
It does:

- Variable prep
- Clever missing values
- Variable selection
- Model selection (few models) & evolutionary param search
- Not: image & time series

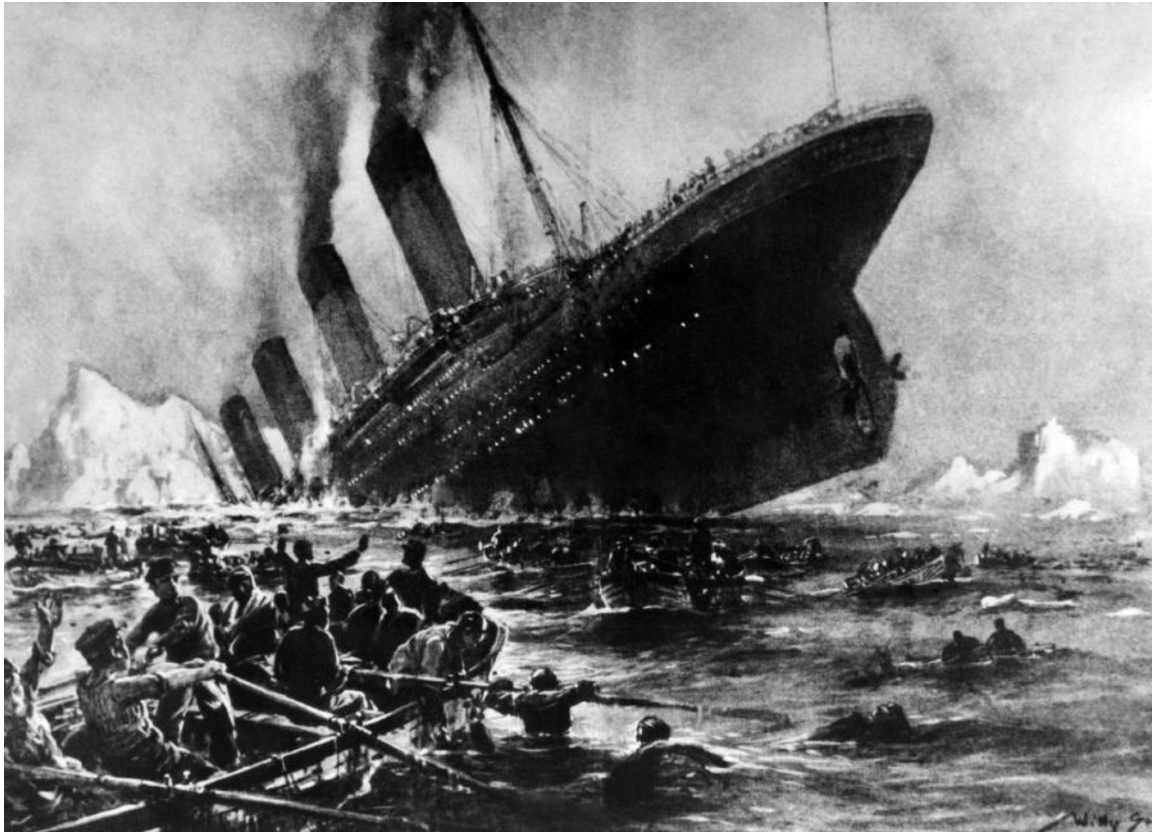


```
pip install xtoy
```

xtoy


NOTHING ON EARTH
COULD COME BETWEEN THEM.
LEONARDO DICAPRIO KATE WINSLET
A JAMES CAMERON FILM
TITANIC

xtoy



Automated machine learning

- 20-80 rule

- Make domain specific ML platform
  - pre-processing
  - cross-validation
  - anomaly detection
  - **Spend extra time on most important & reusable features**

- Add new data

- Underestimate time-to-production

Automated machine learning

- 20-80 rule

- Make domain specific ML platform
  - pre-processing
  - cross-validation
  - anomaly detection
  - **Spend extra time on most important & reusable features**

- Add new data

- Underestimate time-to-production

Conclusion

- Machine learning is just a **tool**
- But can be really **powerful** under the right **circumstances**
- Can you easily create a **feedback loop**?
- **Pluggability** is key
- Don’t try to solve the most **complex** problems!
- Don’t do it when many **strict rules** are already in place
- **Optimizing model** is fun, but usually not the “main gain”
- Never underestimate the **work** required **besides** machine learning
- Build a **framework** (for your company) to handle your typical data

Questions?

https://app.sli.do/event/1vcr8poz/questions

Machine learning

- Does not work when there is a **big policy change**
- **Model should learn to generalize… what does that mean?**
- **Representative data?**
  - When all situations are **unique**, there is **no pattern** to learn!
  - When situations are all the same, then you can just write if/else
- New features can easily be added