H2O Machine Learning Use Cases

Jo-fai (Joe) Chow
Data Scientist
joe@h2o.ai
@matlabulous

satRday #1
MTA TTK, Budapest, Hungary
3rd September, 2016

About Me: Civil Engineer → Data Scientist

- 2005 - 2015
- Water Engineer
  - Consultant for Utilities
  - EngD Research

- 2015 - Present
- Data Scientist
  - Virgin Media (UK)
  - Domino Data Lab (US)
  - H2O.ai (US)

Why? Long story – see bit.ly/joe_h2o_talk2

About this Talk

- Introduction to H2O
- Machine learning use cases from our users

About H2O.ai

- **H2O.ai, the Company**
  - Team: 80 (71 shown)
  - Founded in 2012,
  - HQ: Mountain View, California

- **H2O, the Platform**
  - Open Source (Apache 2.0)
  - R, Python, Scala, Java and Web Interfaces
  - Distributed Algorithms that Scale to Big Data
  - Works with Laptop, Hadoop & Spark

Current Algorithm Overview

satRday Workshop

Statistical Analysis

- Linear Models (GLM)
- Naïve Bayes

Ensembles

- Random Forest
- Distributed Trees
- Gradient Boosting Machine
- R Package - Stacking / Super Learner

Deep Neural Networks

- Multi-layer Feed-Forward Neural Network
- Auto-encoder
- Anomaly Detection
- Deep Features

Clustering

- K-Means

Dimension Reduction

- Principal Component Analysis
- Generalized Low Rank Models

Solvers & Optimization

- Generalized ADMM Solver
- L-BFGS (Quasi Newton Method)
- Ordinary Least-Square Solver
- Stochastic Gradient Descent

Data Munging

- Scalable Data Frames
- Sort, Slice, Log Transform

H2O Interfaces – R, Python & Others

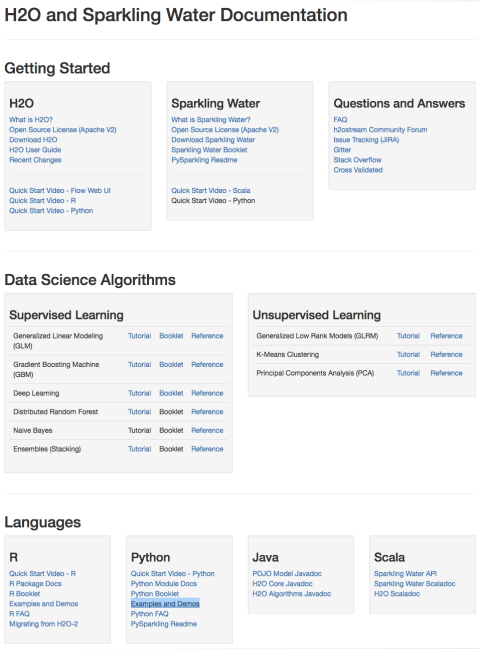
- R

```r
# Load H2O R package
library(h2o)

# Initialize and Connect to H2O
h2o.init()
```

- Python

```python
# Import H2O Python module
import h2o

# Initialize and Connect to H2O
h2o.init()
```

- Resources - docs.h2o.ai



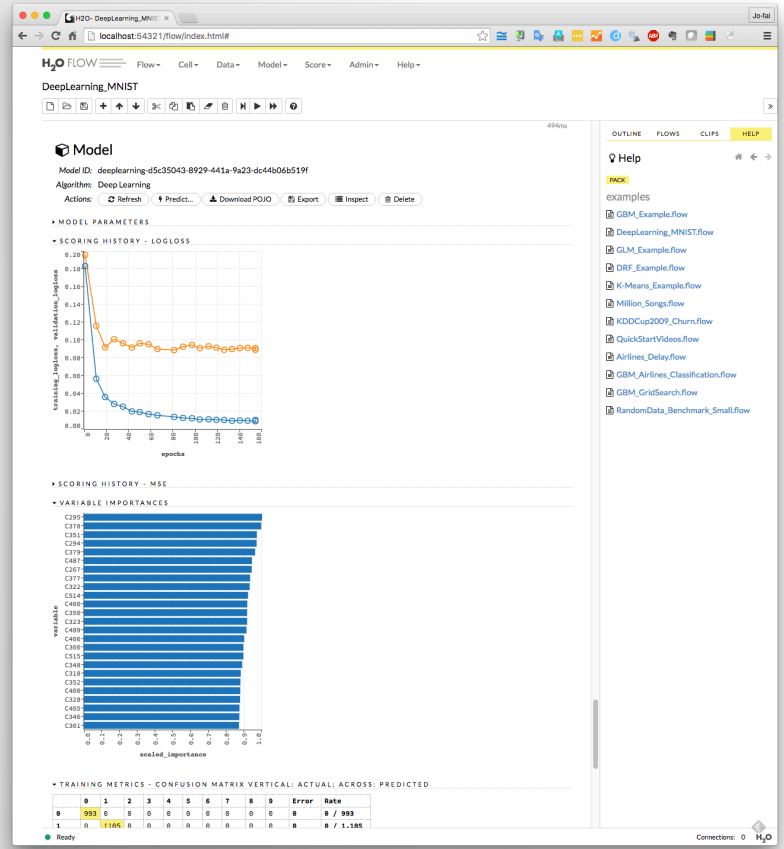
H2O R Code Example

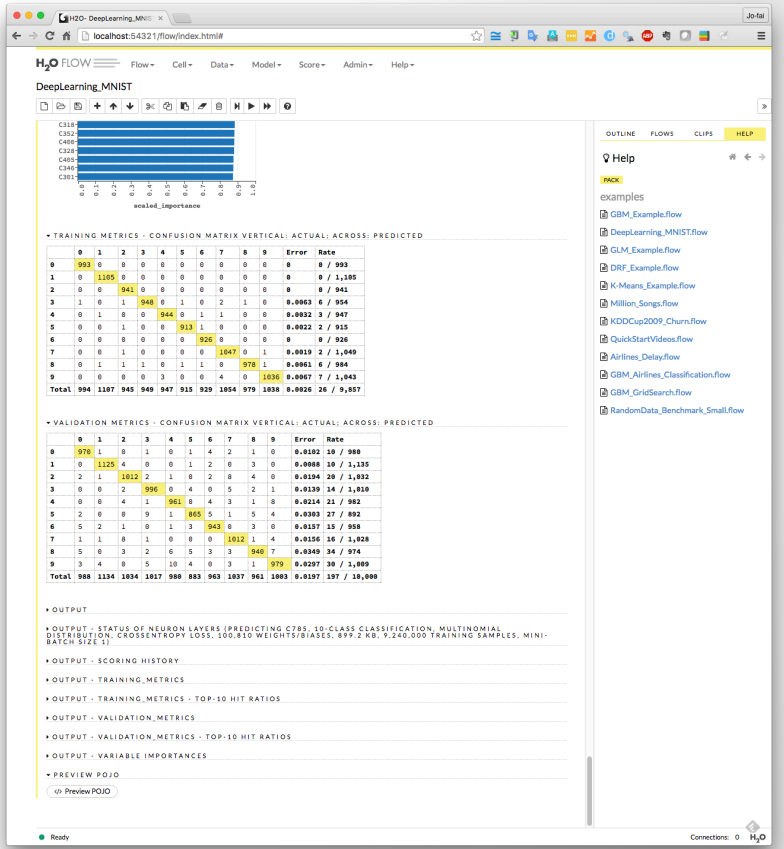
```
42  # Train a GBM with default values
43  model_gbm ← h2o.gbm(x = features,
44                      y = target,
45                      training_frame = h2o_df_boston)
46
47  # First look
48  print(model_gbm)
```

Slide & Code for Workshop:

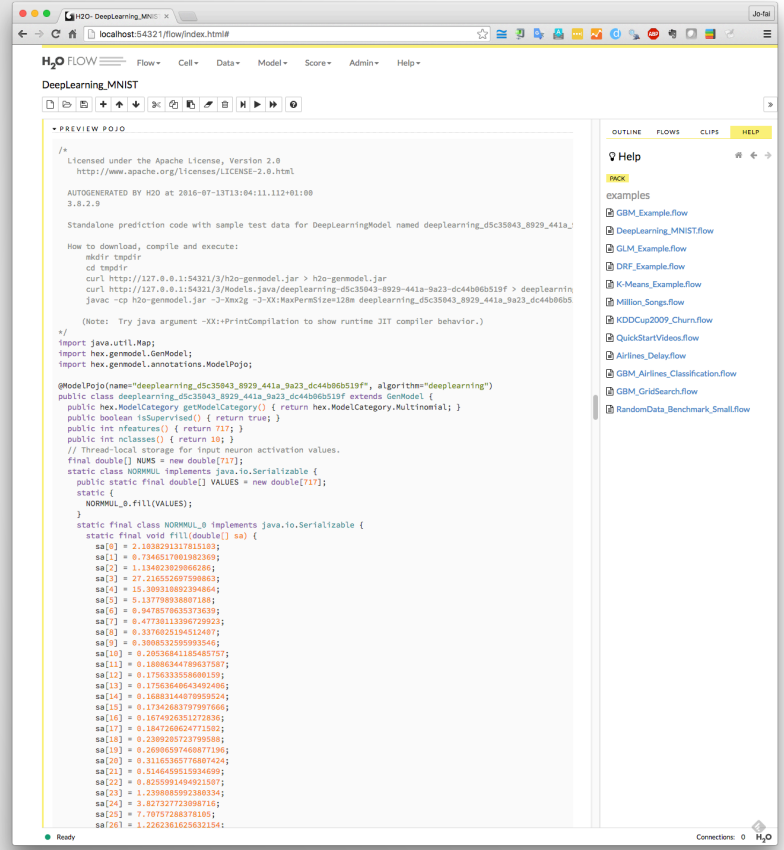
bit.ly/h2o_budapest_workshop

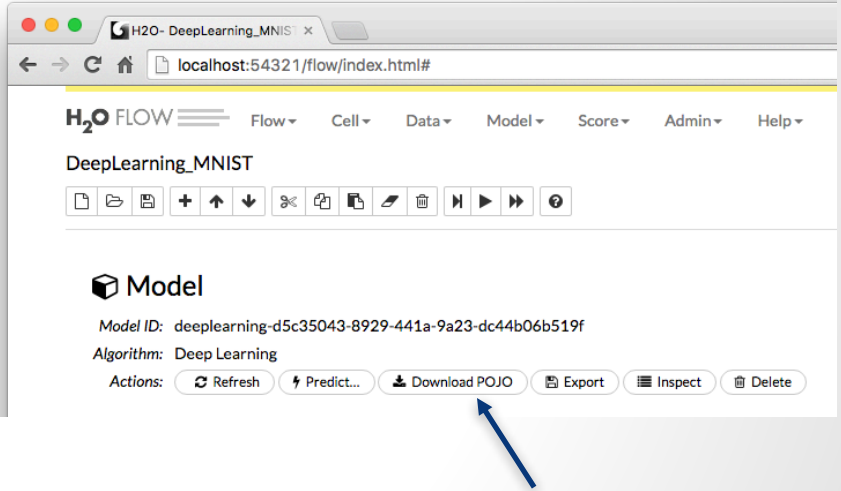
H2O Interfaces – Web (H2O Flow)





Export Plain Old Java Object (POJO)



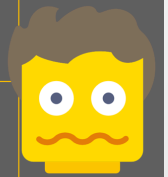


Use Cases

Telenor – Why H2O?

WHAT WERE THE MAIN ASPECTS WE VALUED IN A ML SOLUTION IN 2015?

	R	Spark ML 1.3	Radoop	H_2O
Easy to work with	😐	☹	☺	☺
Cost efficient	☺	☺	☹	☺
Provide the methods that we normally use	☺	😐	☺	☺
Easy-to-use real-time capability	☹	☹	☺	☺
Can utilize our hardware setup	☹	☺	☺	☺

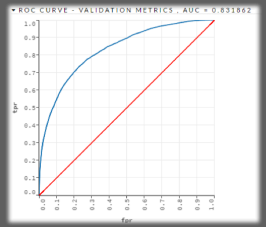


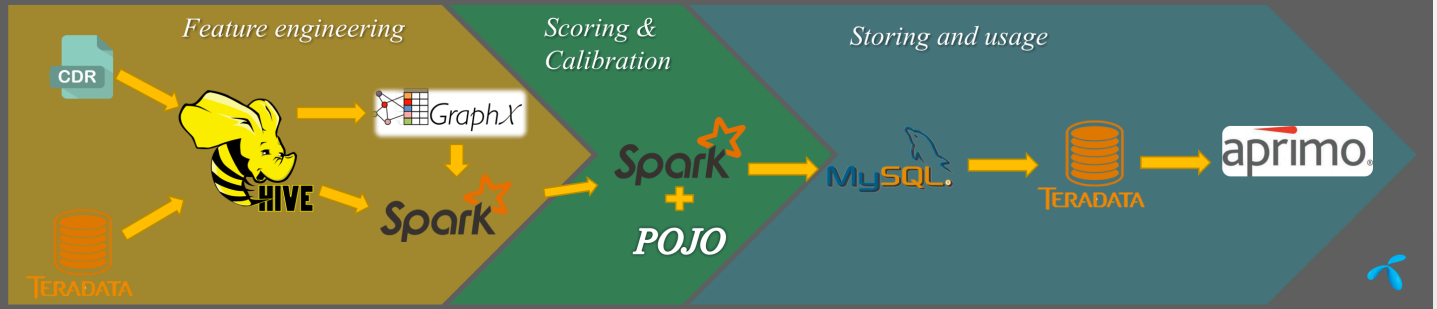
Telenor – Use Case

USE CASE: CHURN MODELLING

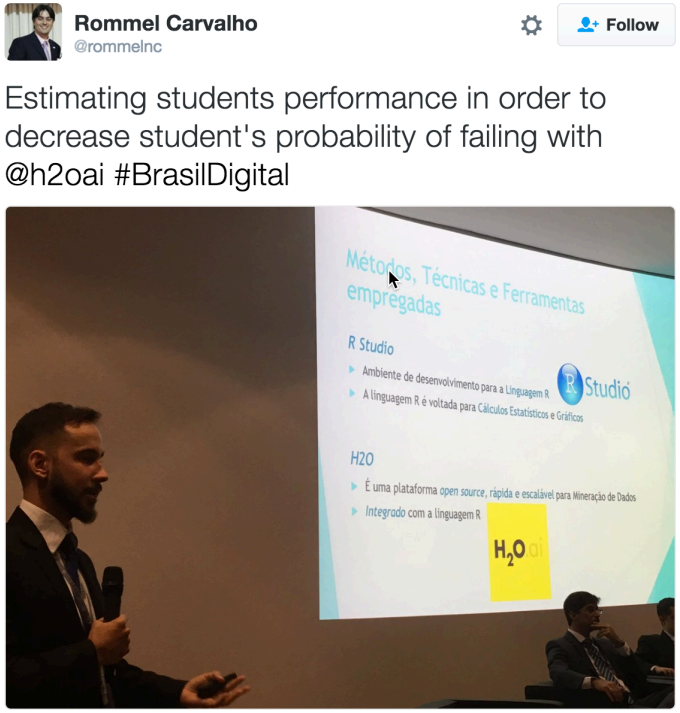
Modelling on 100 GB of data.
- 40 M rows
- 551 columns

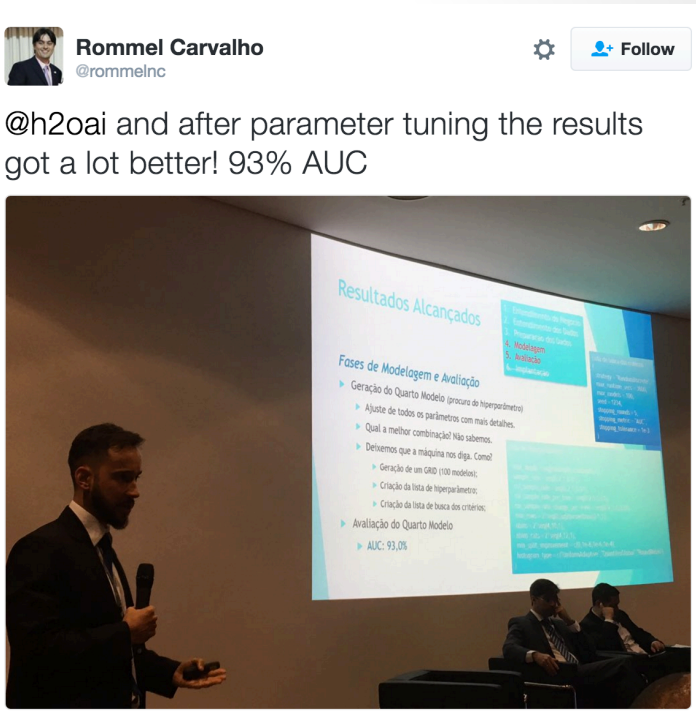
Results:
- GBM
- AUC → 0.83
- LIFT top 1% → 20





H2O in Brazil





R + H2O + Spark at EARL 2016

Vincent Warmerdam | GoDataDriven

I'm a data scientist at GoDataDriven in Amsterdam where I help companies get better at being a company using data. I'm also a preferred Rstudio training partner and co- chair of PyData Amsterdam. I also have a blog that is doing well over at koaning.io.

ML for SparkR: just add water

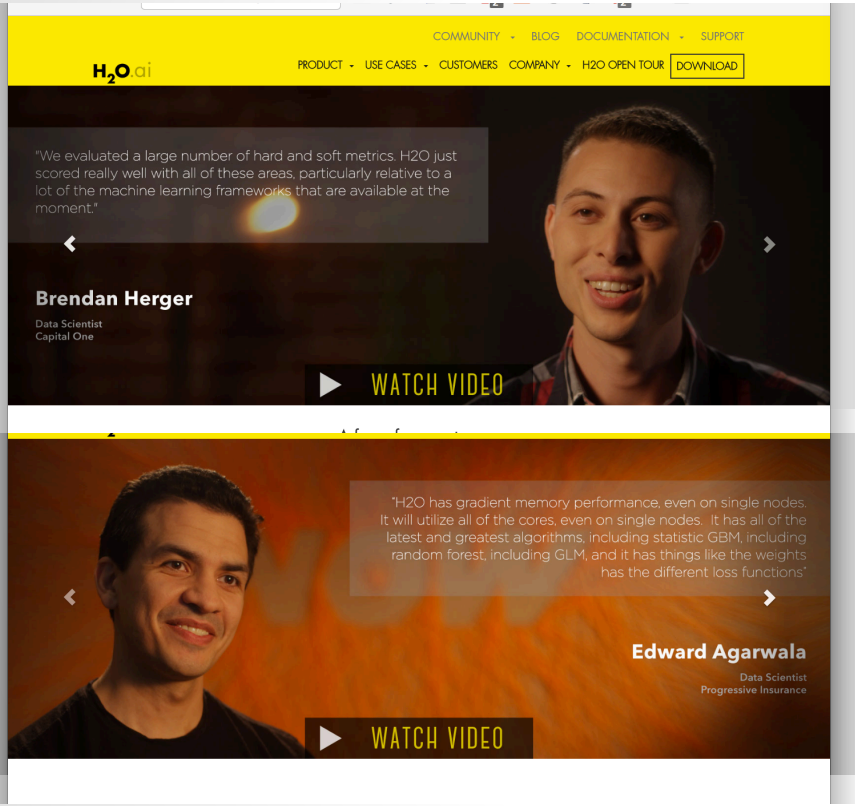
The SparkR project allows more scalable tools for R users that work well with the Rstudio stack as well as the Hadoop stack. The ML capabilities however are somewhat lacking in features.

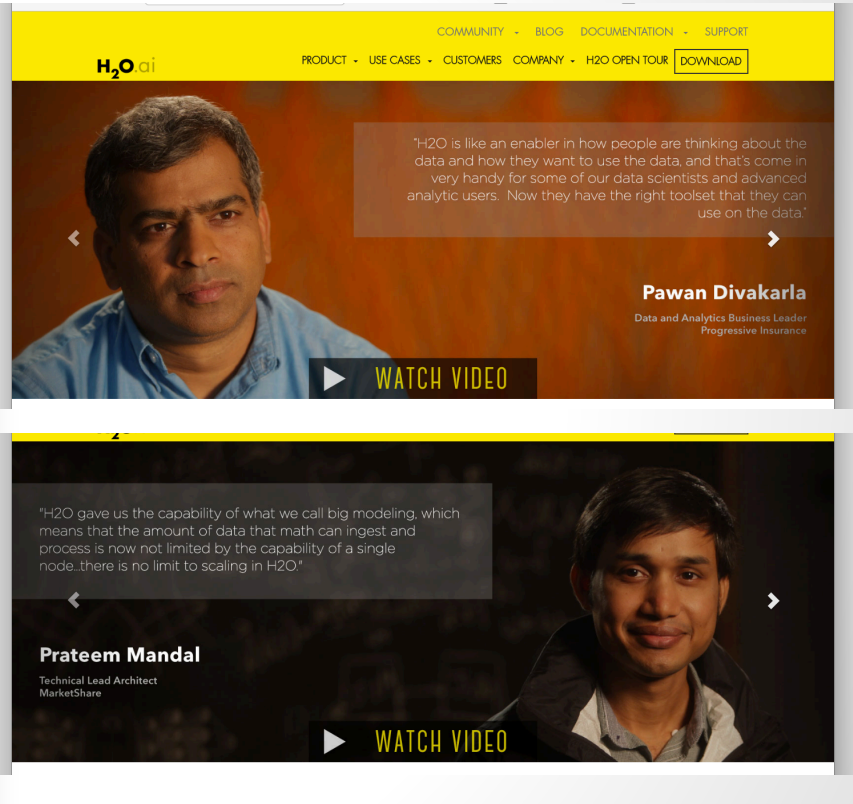
In this small talk I'll demo how to provision SparkR with H20 allowing more algorithms to the R users of Spark. I'll discuss the pros and cons of this approach and demo a few shiny features and use cases of the H20 stack.

Some of the themes I'll discuss;

- ease of provisioning
- works on python stack as well
- your model will be saved as a *.jar which makes it easy for deployment
- grid search is a feature
- deep models on spark

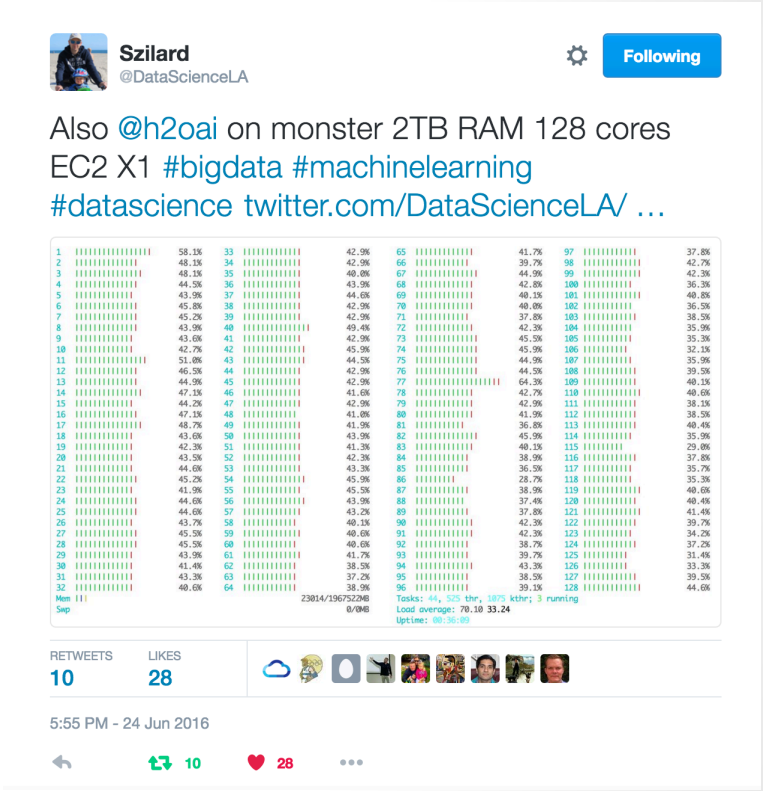
Our Customers





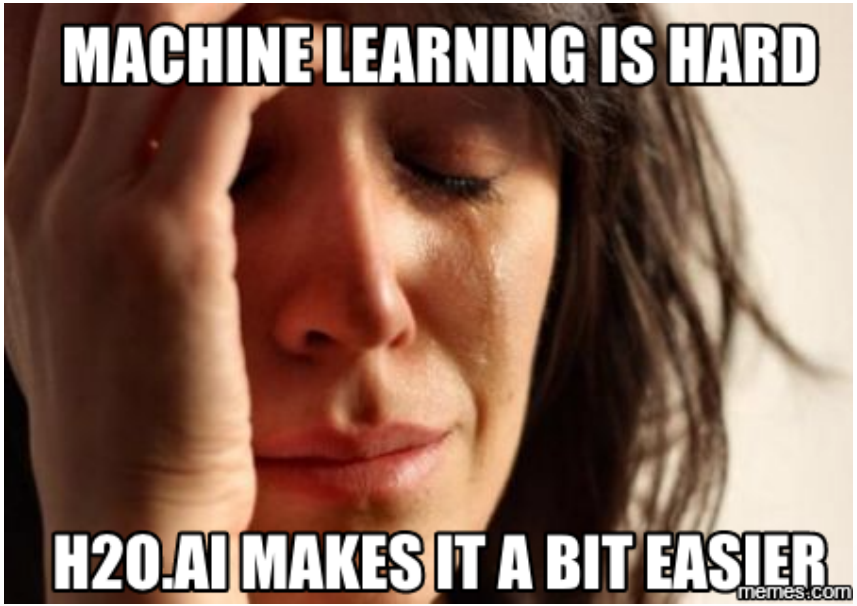
Szilard Pafka

- Budapest DS Meetup
  - Big thanks to Szilard Pafka
    - Intro to ML with H2O
  - Link (video)
  - Link (Slides)



Szilard – Why H2O?

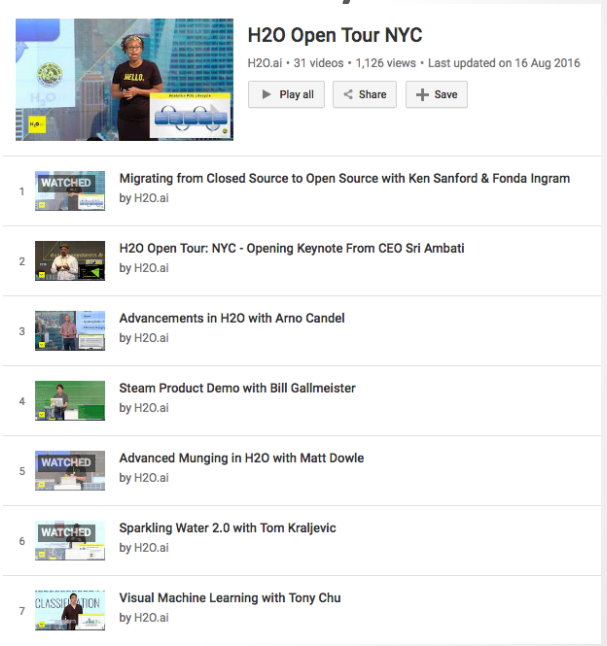
- Szilard's Summary Slide



H2O is Evolving

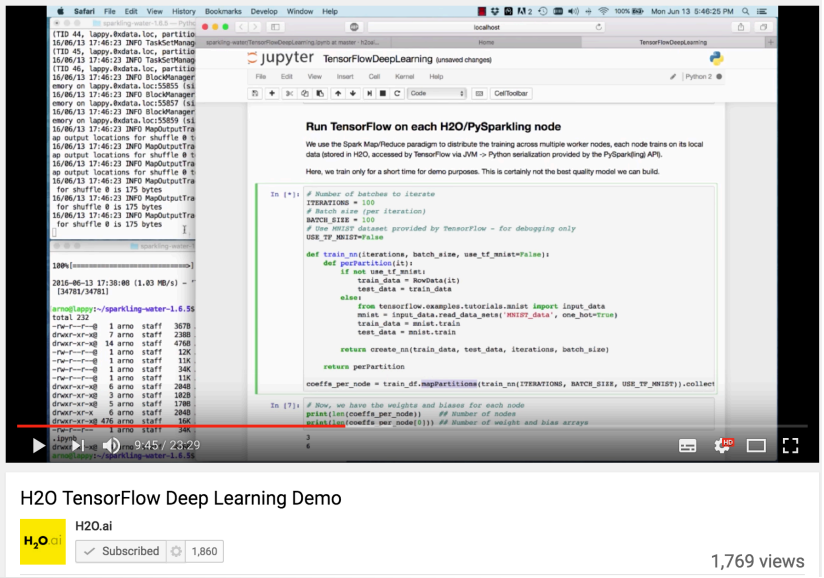
- Advanced data munging
- Visual ML
- Deep Water
  - H2O + TensorFlow, mxnet, Theano, Caffe …
  - GPU
- Steam

- YouTube Playlist

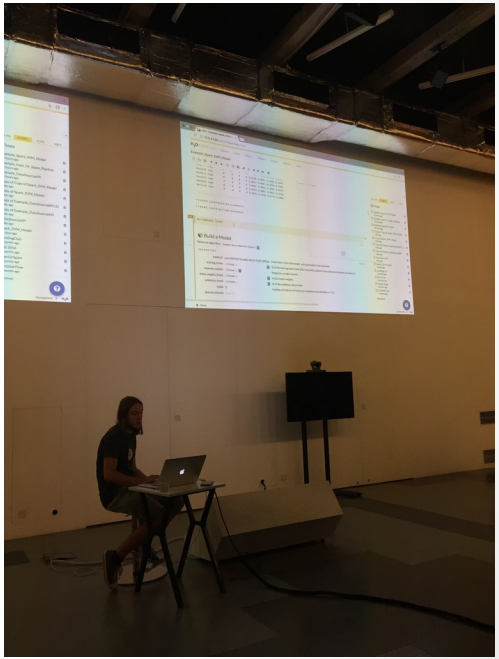


H2O Integration with Other Libraries

- H2O + TensorFlow demo



- H2O + MLLib SVM demo



We're Hiring!

- Check it out
- www.h2o.ai/careers/

Open Positions

Location	Position
Mountain View, CA	UI Engineers
Mountain View, CA	Algorithm Engineers
Mountain View, CA	Solutions Architect, Data Engineering
Mountain View, CA	Distributed Systems Platform Engineer
Mountain View, CA	Customer Support Manager
Mountain View, CA	Quality Engineer
Multiple	Program Manager
Multiple	Solutions Data Scientist
Multiple	Data Journalist

To apply, send your resume to careers@h2o.ai

Thanks!

- Sponsors & Organizers
  - satRdays
  - Budapest Data Science Meetup

- Contact
  - joe@h2o.ai
  - @matlabulous
  - github.com/woobe

- Resources
  - docs.h2o.ai

- Slides and Code
  - github.com/h2oai/h2o-meetups