# Machine Learning with H2O


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

Budapest Data Science Meetup
Prezi House of Ideas
2nd 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

# I love Prezi!


Share Embed Like Public

## Quantifying Green Values

No description

by Jo-fai Chow on 28 June 2013 • 998 1 Tweet

**Comments** (0)

**Popular presentations**

See more popular or the latest prezis

My favorite presentation during my EngD time.

# 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

# Scientific Advisory Council

## Dr. Trevor Hastie

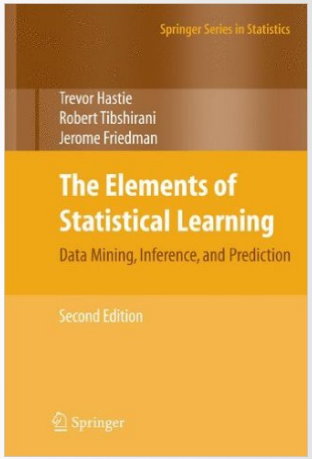
- John A. Overdeck Professor of Mathematics, Stanford University
- PhD in Statistics, Stanford University
- Co-author, *The Elements of Statistical Learning: Prediction, Inference and Data Mining*
- Co-author with John Chambers, *Statistical Models in S*
- Co-author, *Generalized Additive Models*

## Dr. Robert Tibshirani

- Professor of Statistics and Health Research and Policy, Stanford University
- PhD in Statistics, Stanford University
- Co-author, *The Elements of Statistical Learning: Prediction, Inference and Data Mining*
- Author, *Regression Shrinkage and Selection via the Lasso*
- Co-author, *An Introduction to the Bootstrap*

## Dr. Steven Boyd

- Professor of Electrical Engineering and Computer Science, Stanford University
- PhD in Electrical Engineering and Computer Science, UC Berkeley
- Co-author, *Distributed Optimization and Statistical Learning via the Alternating Direction Method of Multipliers*
- Co-author, *Linear Matrix Inequalities in System and Control Theory*
- Co-author, *Convex Optimization*



# Current Algorithm Overview

### 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

See bit.ly/joe_h2o_talk4

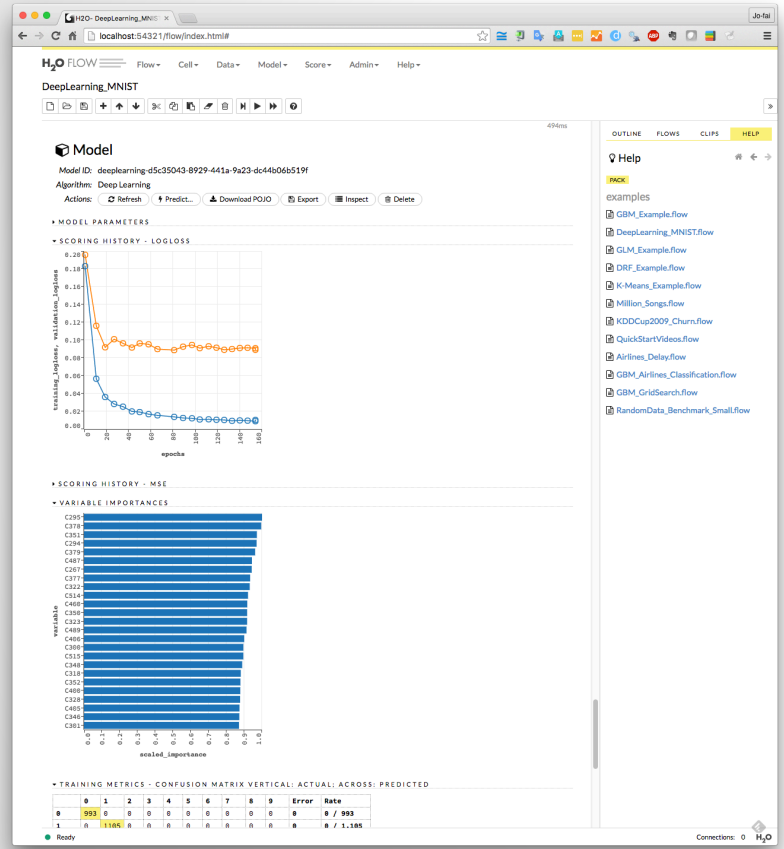
### Solvers & Optimization

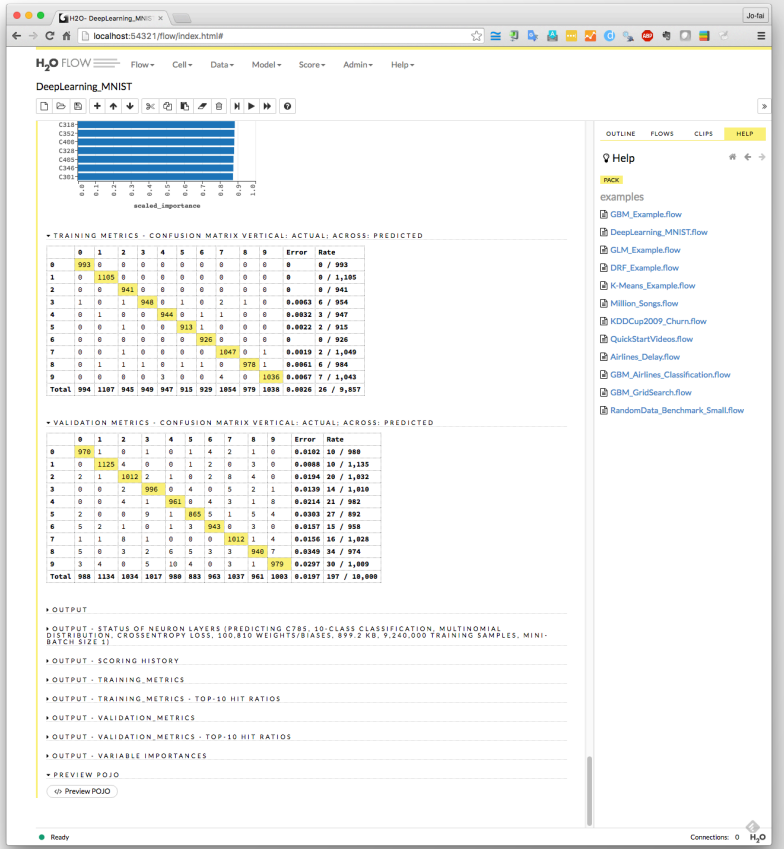
- Generalized ADMM Solver
- L-BFGS (Quasi Newton Method)
- Ordinary Least-Square Solver
- Stochastic Gradient Descent

### Data Munging

- Scalable Data Frames
- Sort, Slice, Log Transform

# H2O Interfaces – Web (H2O Flow)



# H2O Interfaces – R, Python & Others

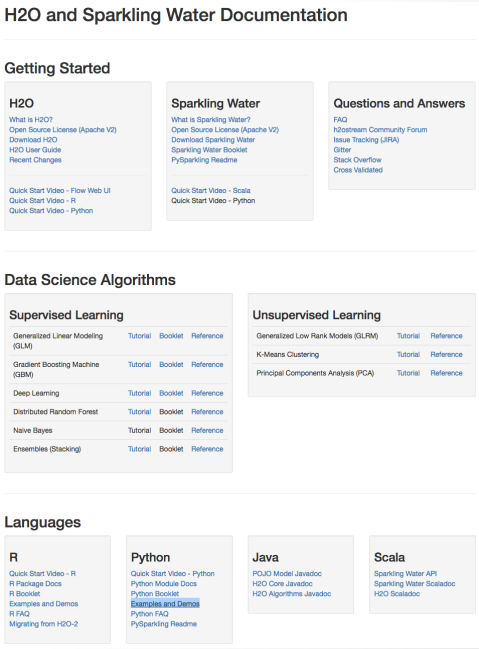
- R

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

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

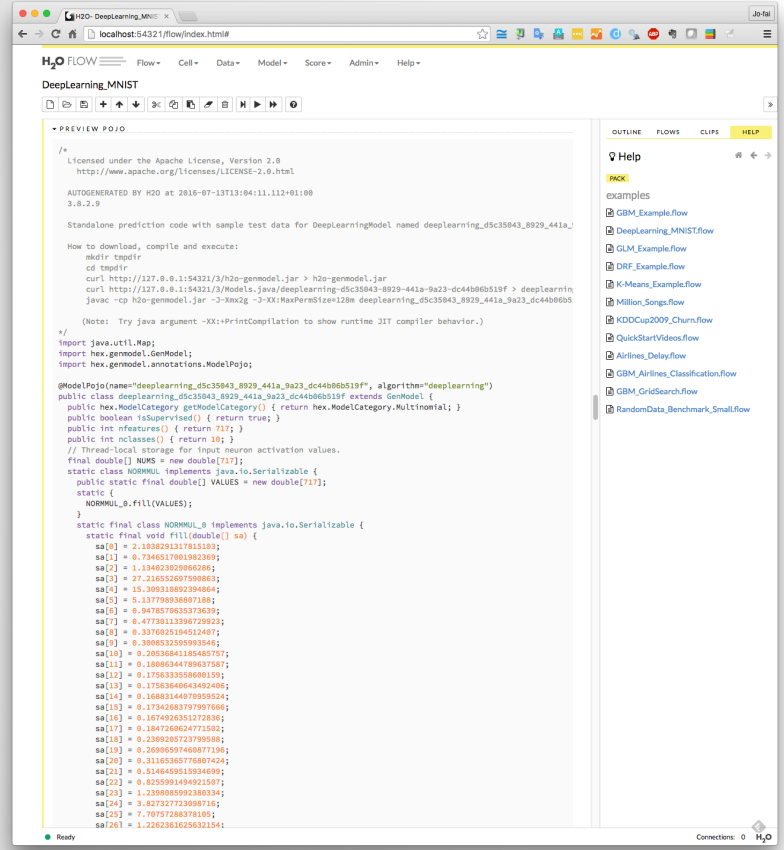
- Python

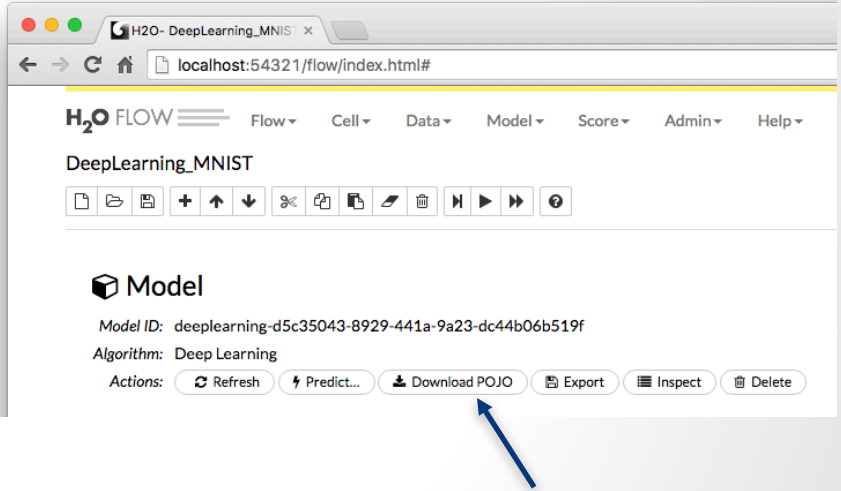
```python
# Import H2O Python module
import h2o

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

- Resources - docs.h2o.ai



# Export Plain Old Java Object (POJO)

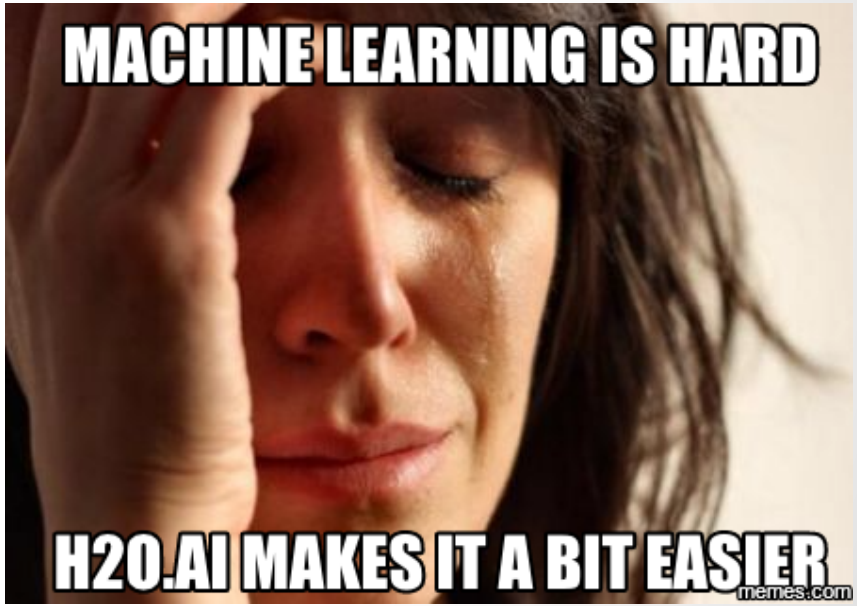


# Why H2O?

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

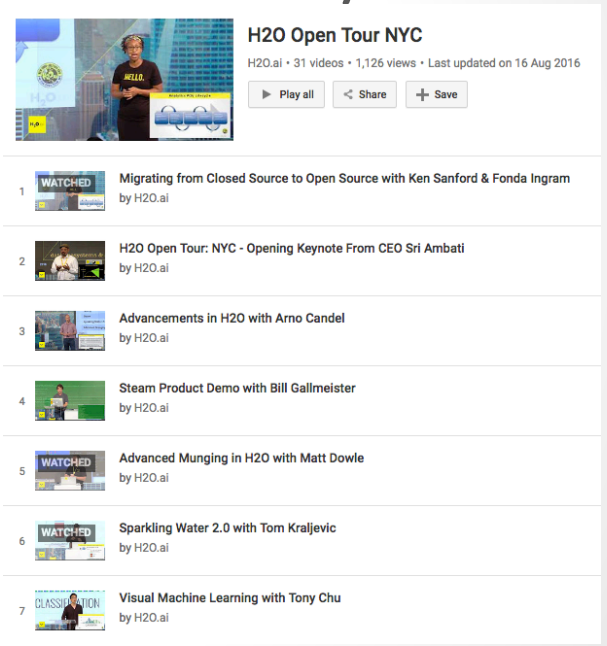
- Szilard's Summary Slide



# H2O is Evolving

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

- YouTube Playlist



# H2O at satRday #1

- **Hands-on Workshop (0800)**
  - H2O R package
  - Scalable machine learning
  - Model training, tuning and stacking

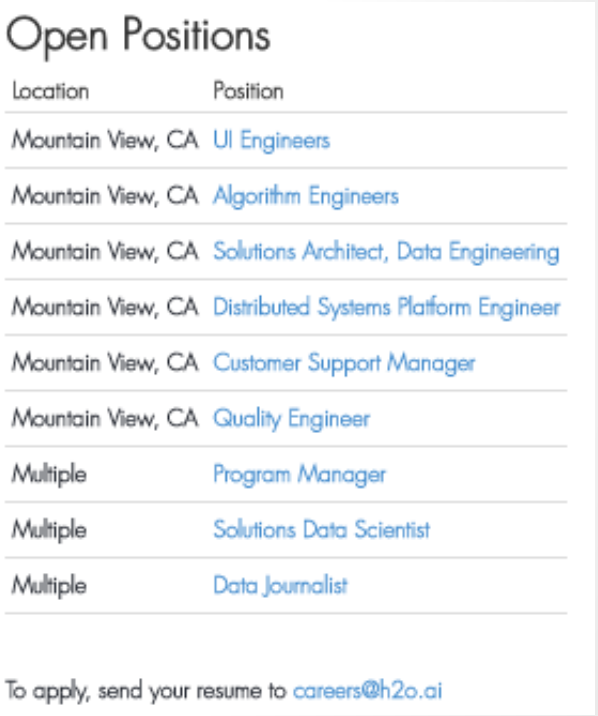
- **Machine Learning Use Cases (1600)**
  - How others use H2O for their projects

# Three More H2O Talks

- Zsolt Toth
  - Rapid Miner + H2O Integration (20 mins)
- Norbert Liki
  - Data Mining with H2O at Telenor (20 mins)
- Jakub (Kuba) Háva
  - Sparkling Water 2.0 (35 mins)

# 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!

- Prezi
- Organizers
  - Budapest DS Meetup
  - satRdays
- Szilard
- H2O.ai

- Contact
  - joe@h2o.ai
  - @matlabulous
  - github.com/woobe
- Resources
  - docs.h2o.ai
- Slides and Code
  - github.com/h2oai/h2o-meetups