

# Deep Water

The logo for H2O.ai is displayed on a yellow square background. It features the text "H2O.ai" in a bold, sans-serif font. The "H2O" is in black, and the ".ai" is in a lighter yellow color.

**H<sub>2</sub>O.ai**

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

Data Science for IoT Meetup  
Barclays Eagle Venture Labs  
21<sup>st</sup> & 24<sup>th</sup> November, 2016

# H2O Deep Learning Is Widely Used

The usage of Hadoop/Big Data tools grew to 39%, up from 29% in 2015 (and 17% in 2014), driven by Apache Spark, MLlib (Spark Machine Learning Library) and H2O.

See also

- KDnuggets interview with Spark Creator Matei Zaharia
- KDnuggets interview with Arno Candel, H2O.ai on How to Quick Start Deep Learning with H2O

<http://www.kdnuggets.com>

H2O and TensorFlow are tied



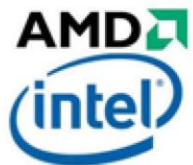
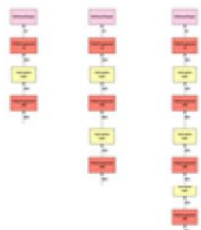
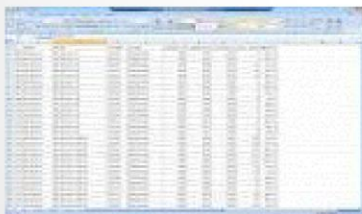
Deep Learning Tools & Platforms: Caffe (63)	0%
Cuda-convnet (22)	0%
Convnet.js (3)	0%
carch (2)	0%
Deeplearning4j (46)	0%
H2O (190)	1%
MATLAB Deep Learning Toolbox (57)	0%
Microsoft CNTK (25)	0%
mxnet (16)	0%
Nervana (2)	0%
Tensorflow (190)	1%
Theano ecosystem including Keras, Lasagne, Pylearn2 (140)	1%
Torch (28)	0%
Veles (2)	0%
Other Deep Learning Tools (104)	1%

usage of Deep Learning tools in past year

# Deep Water opens the Floodgates for state-of-the-art Deep Learning

## H2O Deep Learning: simple multi-layer neural networks

1-5 layers  
MBs/GBs of data

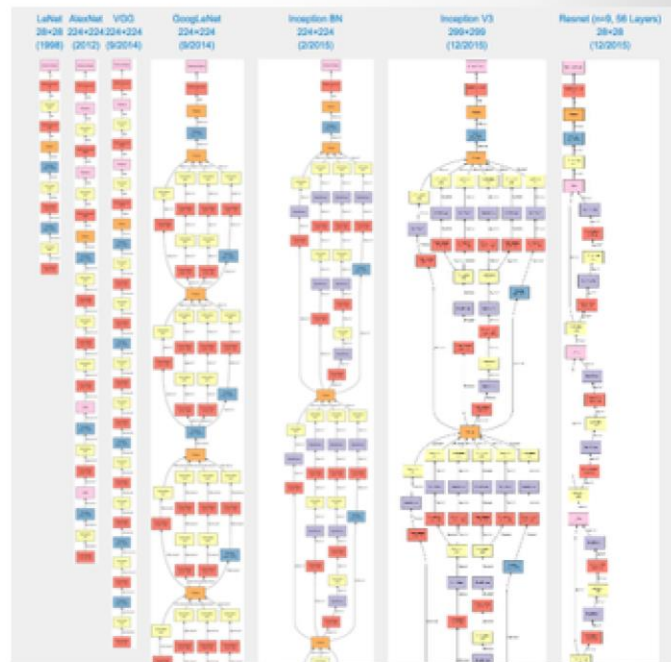
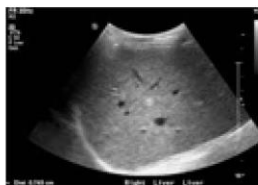


Limited to business analytics,  
statistical models (CSV data)

H<sub>2</sub>O.ai

## Deep Water: deep complex networks




5-1000 layers  
GBs/TBs of data



Large networks for big data  
(e.g. image 1000x1000x3 -> 3m inputs)

# Deep Water: Best Open-Source Deep Learning

## Enterprise Deep Learning for Business Transformation

Deep Water = THE Deep Learning Platform	H2O integrates the top open-source DL tools	
Native GPU support	 is up to 100x faster than 	
Enterprise Ready	Easy to train and deploy, interactive, scalable, etc. Flow, R, Python, Spark/Scala, Java, REST, POJO, <b>Steam</b>	
New Big Data Use Cases (previously impossible or difficult in H2O)	<b>Image</b> - social media, manufacturing, healthcare, ... <b>Video</b> - UX/UI, security, automotive, social media, ... <b>Sound</b> - automotive, security, call centers, healthcare, ... <b>Text</b> - NLP, sentiment, security, finance, fraud, ... <b>Time Series</b> - security, IoT, finance, e-commerce, ...	

H2O FLOW

FlowCellDataModelScoreAdminHelp

Untitled Flow

CS

Expression...

Using Flow to train Deep Water Model

Deep Learning...  
Deep Water...  
Distributed Random Forest...  
Gradient Boosting Method...  
Generalized Linear Modeling...  
Generalized Low Rank Modeling...  
K-means...  
Naive Bayes...  
Principal Components Analysis...  
  
List All Models  
List Grid Search Results  
Import Model...  
Export Model...

Ready

Connections: 0

H2O

H2O FLOW

Flow

Cell

Data

Model

Score

Admin

Help

Untitled Flow



CS

Expression...

Using Flow to train Deep Water Model

Deep Learning...

Deep Water...

Distributed Random Forest...

Gradient Boosting Method...

Generalized Linear Modeling...

Generalized Low Rank Modeling...

K-means...

Naive Bayes...

Principal Components Analysis...

List All Models

List Grid Search Results

Import Model...

Export Model...

Ready

Connections: 0

H2O



# Same H2O R/Python Interface

To build a LeNet image classification model in H2O, simply specify network = "lenet":

```
model <- h2o.deepwater(x=path, y=response,  
                       training_frame=df, epochs=50,  
                       learning_rate=1e-3, network = "lenet")
```

model

|=====| 100%

Model Details:

=====

H2OMultinomialModel: deepwater

Model ID: DeepWater\_model\_R\_1477378862430\_2

Status of Deep Learning Model: lenet, 1.6 MB, predicting C2, 3-class classification, 14,336 training samples, mini-batch size 32

	input_neurons	rate	momentum
1	2352	0.000986	0.990000

H2OMultinomialMetrics: deepwater

\*\* Reported on training data. \*\*

\*\* Metrics reported on full training frame \*\*

Training Set Metrics:

=====

Extract training frame with `h2o.getFrame("cat\_dog\_mouse.hex\_sid\_95f8\_1")`

MSE: (Extract with `h2o.mse`) 0.131072

RMSE: (Extract with `h2o.rmse`) 0.3620386

Logloss: (Extract with `h2o.logloss`) 0.4176429

## Roadmap for Deep Water (Q4 2016):



**Finish TensorFlow integration (C++/Python/Java):**  
**Package Python on the backend to create trainable graphs**



**Finish Caffe integration (pure C++/Java):**  
**Optimized Multi-GPU training (NVIDIA NCCL)**

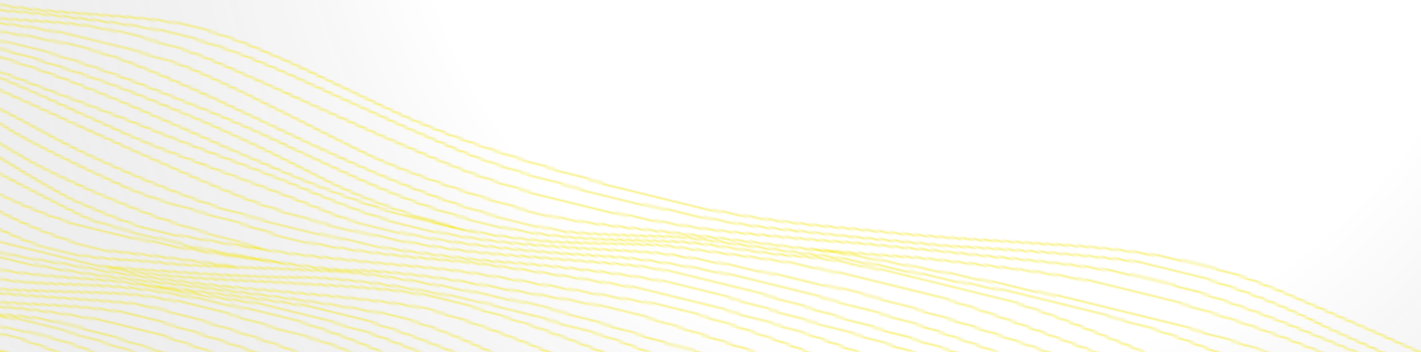


**Add multi-GPU support for mxnet**



**Add more capabilities to H2O Deep Water:**  
**Text/NLP, Time Series, LSTM, AutoEncoder,**  
**Feature Extraction, Input/Output shape mapping, etc.**

# Deep Water Demo





# Deep Water Demo

- H2O + mxnet
  - Datasets:
    - Cat / Dog / Mouse
    - Iris
  - mxnet GPU backend
  - Training a LeNet (CNN) model
  - Using random grid search for hyper-parameters optimization
- Code and Data
  - [bit.ly/h2o\\_warsaw\\_1](https://bit.ly/h2o_warsaw_1)
  - [github.com/h2oai/deepwater](https://github.com/h2oai/deepwater)

# Code and References

## Python/R Jupyter Notebooks

Check out a sample of cool Deep Learning [Jupyter notebooks](#)!

## PreRelease Downloads

For the following system dependencies, we provide recent builds for your convenience.

- Ubuntu 16.04 LTS
- Latest NVIDIA Display driver
- CUDA 8 (latest available) in `/usr/local/cuda`
- CUDNN 5 (inside of `lib` and include directories in `/usr/local/cuda/`)

In the future, we'll have more pre-built jars for more OS/CUDA combinations.

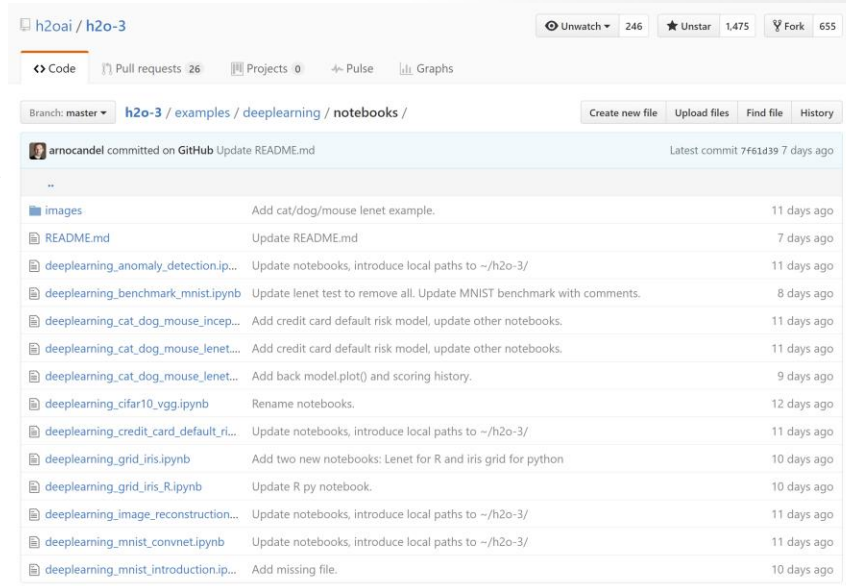
- Required to run Jupyter notebook: `H2O Deep Water enabled Python module` -- install via `pip install <file>`
- To build custom networks: `Matching MXNet Python egg` -- install via `easy_install <file>`
- To run from Flow only: `H2O Standalone h2o.jar` -- launch via `java -jar h2o.jar`

If you are interested in running H2O Deep Water on a different infrastructure, see the DIY build instructions below

## PreRelease Amazon AWS Image

For your convenience, here's a pre-built image for Amazon's EC2 environment, based off our recent [H2O Open Tour Hands-On Deep Water workshop](#) (recording coming soon).

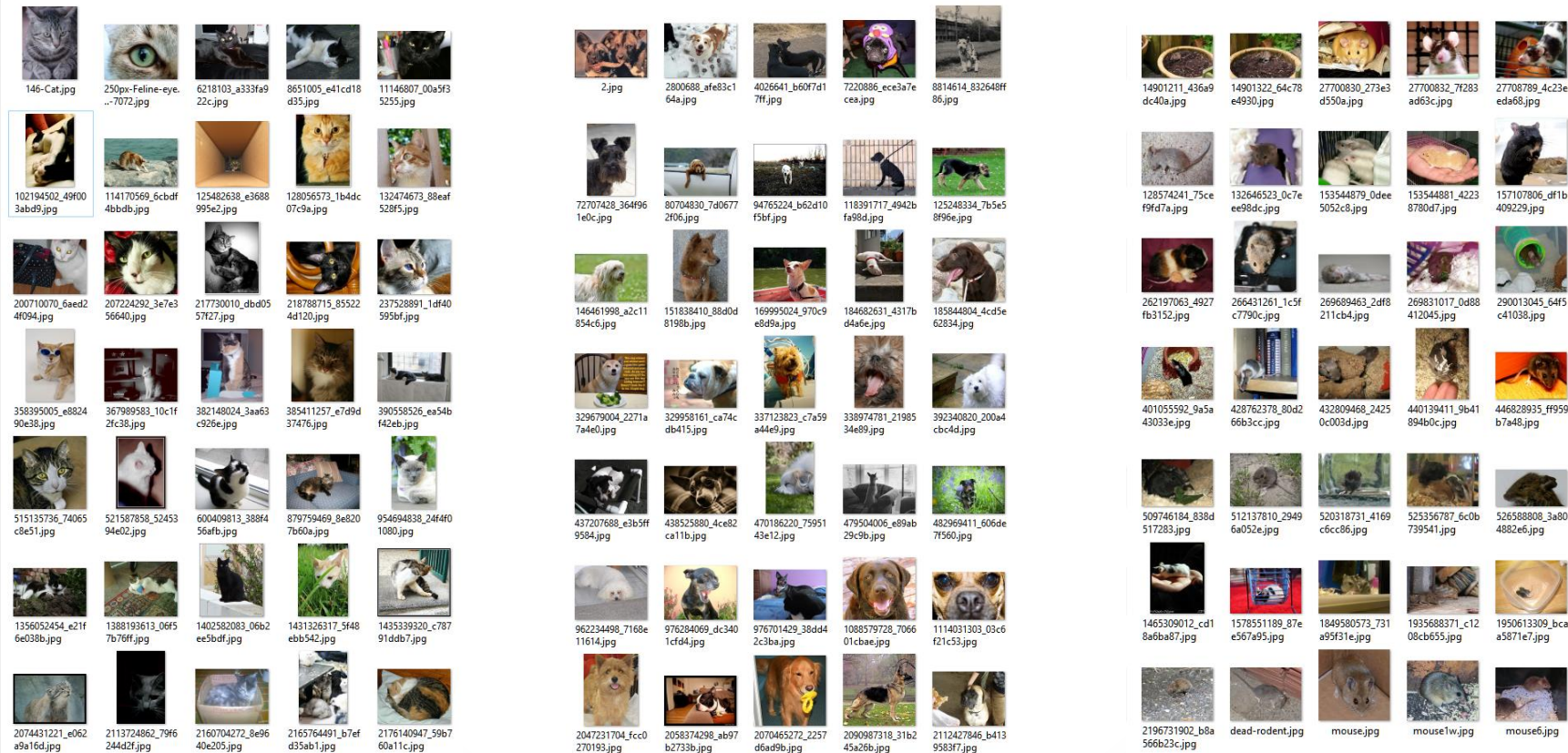
- AMI ID: `ami-d32f70c4`
- AMI Name: `deepwater-dallas-v3`
- Recommended instance types: `g2.2xlarge` or `p2.xlarge`
- After launching the instance, you can connect to port 8888 (Jupyter Notebook) or port 54321 (H2O Flow).



The screenshot shows the GitHub repository page for `h2oai/h2o-3`. The repository has 246 pull requests, 1,475 stars, and 655 forks. The current branch is `master`. The file path `h2o-3 / examples / deeplearning / notebooks /` is selected. The commit history shows a series of updates to the notebooks, including adding cat/dog/mouse lenet examples, updating README.md, and adding various deeplearning notebooks like `deeplearning_anomaly_detection.ipynb`, `deeplearning_benchmark_mnist.ipynb`, `deeplearning_cat_dog_mouse_incep...`, `deeplearning_cat_dog_mouse_lenet...`, `deeplearning_cat_dog_mouse_lenet...`, `deeplearning_cifar10_vgg.ipynb`, `deeplearning_credit_card_default_rl...`, `deeplearning_grid_iris.ipynb`, `deeplearning_grid_iris_R.ipynb`, `deeplearning_image_reconstruction...`, `deeplearning_mnist_convnet.ipynb`, and `deeplearning_mnist_introduction.ip...`.

[github.com/h2oai/deepwater](https://github.com/h2oai/deepwater)

# Data – Cat/Dog/Mouse Images



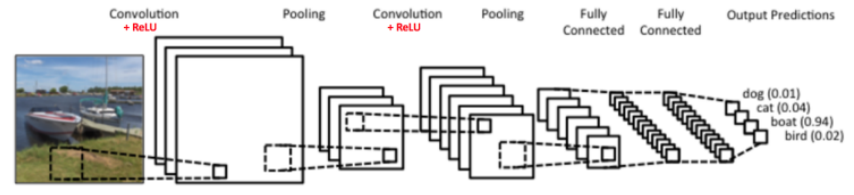
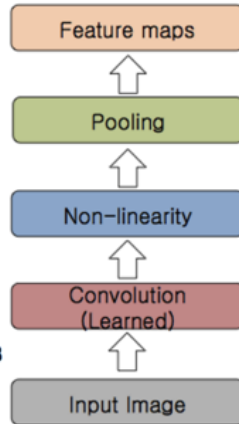
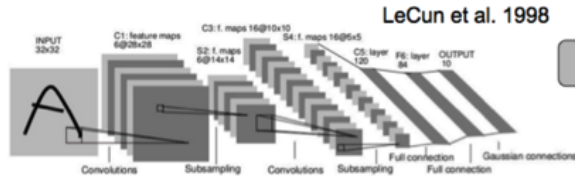
# Data - CSV

	A	B
1	bigdata/laptop/deepwater/imagenet/cat/102194502_49f003abd9.jpg	cat
2	bigdata/laptop/deepwater/imagenet/cat/11146807_00a5f35255.jpg	cat
3	bigdata/laptop/deepwater/imagenet/cat/1140846215_70e326f868.jpg	cat
4	bigdata/laptop/deepwater/imagenet/cat/114170569_6cbdf4bbdb.jpg	cat
5	bigdata/laptop/deepwater/imagenet/cat/1217664848_de4c7fc296.jpg	cat
6	bigdata/laptop/deepwater/imagenet/cat/1241603780_5e8c8f1ced.jpg	cat
7	bigdata/laptop/deepwater/imagenet/cat/1241612072_27ececbbdef.jpg	cat
8	bigdata/laptop/deepwater/imagenet/cat/1241613138_ef1d82973f.jpg	cat
9	bigdata/laptop/deepwater/imagenet/cat/1244562192_35becd66bd.jpg	cat
10	bigdata/laptop/deepwater/imagenet/cat/125482638_e3688995e2.jpg	cat
11	bigdata/laptop/deepwater/imagenet/cat/128056573_1b4dc07c9a.jpg	cat
12	bigdata/laptop/deepwater/imagenet/cat/12945197_75e607e355.jpg	cat
13	bigdata/laptop/deepwater/imagenet/cat/132474673_88eaf528f5.jpg	cat
14	bigdata/laptop/deepwater/imagenet/cat/1350530984_ecf3039cf0.jpg	cat
15	bigdata/laptop/deepwater/imagenet/cat/1351606235_c9fbeb634.jpg	cat
16	bigdata/laptop/deepwater/imagenet/cat/1356052454_e21f6e038b.jpg	cat
17	bigdata/laptop/deepwater/imagenet/cat/1388193613_06f57b76ff.jpg	cat

# LeNet – Convolutional Neural Network

## Convolutional Network

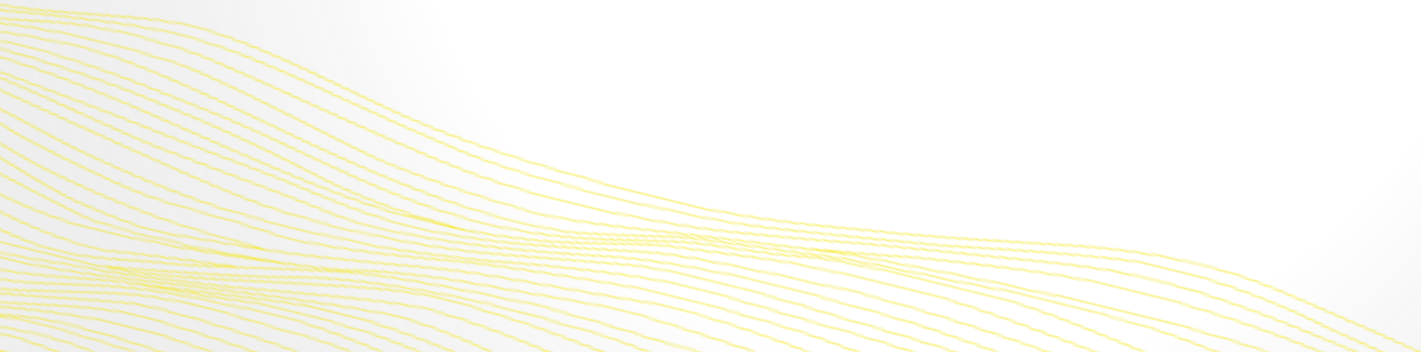
- Feed-forward:
  - Convolve input
  - Non-linearity (rectified linear)
  - Pooling (local max)
- Supervised
- Train convolutional filters by back-propagating classification error



Slide: R. Fergus



# Deep Water Demo



## Current Contributors (more H2O.ai folks joining soon)



Fabrizio Milo



Cyprien Noel



Qiang Kou



Arno Candel



Caffe



H<sub>2</sub>O.ai



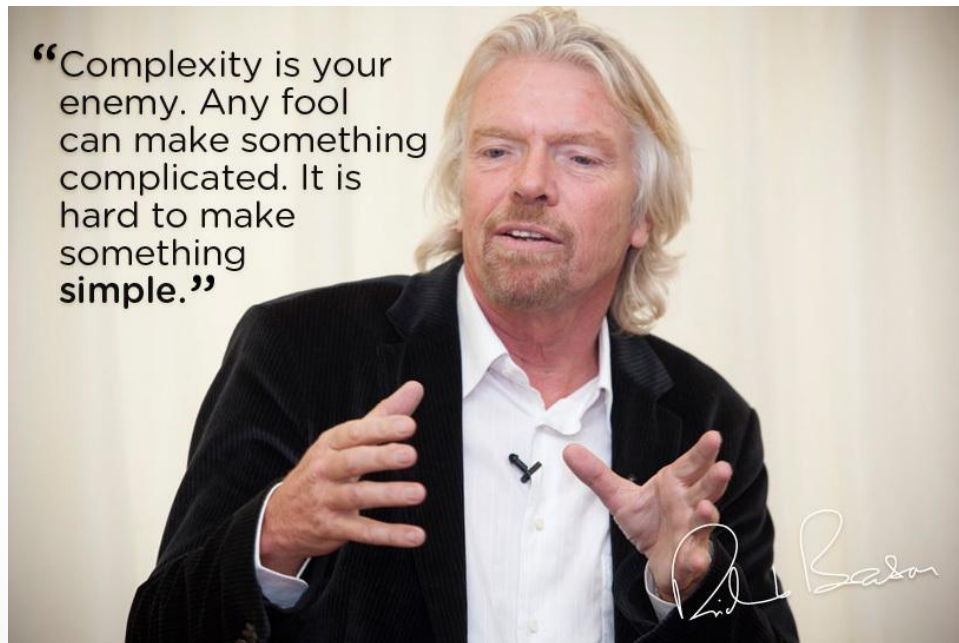
This repository

Search

h2oai / deepwater

# H2O's Mission

Making Machine Learning Accessible to Everyone



*Photo credit: Virgin Media*

# Thanks!

- Organisers & Contributors
  - Ajit Jaokar
  - Sibanjean Das
- Key Resources
  - docs.h2o.ai
  - github.com/h2oai/h2o-meetups
- Slides & Code
  - bit.ly/h2o\_iot\_workshop1
- Contact
  - joe@h2o.ai
  - @matlabulous
  - github.com/woobe