



# Accelerating AI Deployment with H<sub>2</sub>O Driverless AI on IBM Power9

*Jo-fai (Joe) Chow, Data Science Evangelist*

**H2O.ai**

joe@h2o.ai | @h2oai

# OpenPOWER Summit Europe

RAI Centre | Amsterdam  
October 3-4, 2018



Join the Conversation #OpenPOWERSummit

# H2O.ai Overview

Company	Founded in Silicon Valley in 2012 Funded: \$75M Investors: Wells Fargo, NVIDIA, Nexus Ventures, Paxion Ventures
Products	<ul style="list-style-type: none"><li>• H2O Open Source Machine Learning (14,000 organizations)</li><li>• H2O Driverless AI – Automatic Machine Learning</li></ul>
Leadership	Leader in Gartner MQ Machine Learning and Data Science Platform
Team	120 AI experts (Kaggle Grandmasters, Distributed Computing, Visualization)
Global	Mountain View, London, Prague, India



# A Growing Customer Base



*"H2O.ai's reference customers gave it the highest overall score for sales relationship and overall service and support" - Gartner MQ 2018*

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

# Growing Worldwide Open Source Community



14,000 Companies using H<sub>2</sub>O



155,000 data scientists

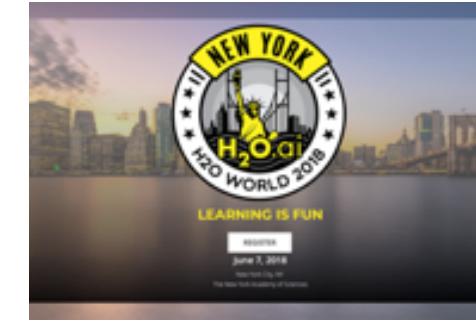
222 OF FORTUNE  
500

LOVE H<sub>2</sub>O

8 OF TOP 10  
BANKS

7 OF TOP 10  
INSURANCE COMPANIES

4 OF TOP 10  
HEALTHCARE COMPANIES



H2O World  
NYC, London, SF  
Thousands attending live and online



116K Meet up Members

# H2O.ai is a **Leader** in the 2018 Gartner Data Science and Machine Learning Platforms Magic Quadrant

- Technology leader with most completeness of vision
- Recognized for the mindshare, partner network and status as a **quasi-industry standard** for machine learning and AI
- H2O.ai customers gave the highest overall score among all the vendors for sales relationship and account management, customer support (onboarding, troubleshooting, etc.) and overall service and support

Figure 1. Magic Quadrant for Data Science and Machine-Learning Platforms



Get the  
Gartner  
Magic  
Quadrant  
[here](#)

# Partner Ecosystem

**SCAN**  
computers



PENGUIN  
COMPUTING

**AMAX**

SUPERMICRO



Google  
Cloud Platform

Microsoft Azure



IBM Cloud

NVIDIA

IBM

TRACE3  
World Wide Technology, Inc.

TechData

accenture

Capgemini

snowflake

Hortonworks

cloudera

MINIO

MAPR

HW Vendors

Cloud Providers

Strategic  
Partners

Value Added  
Resellers

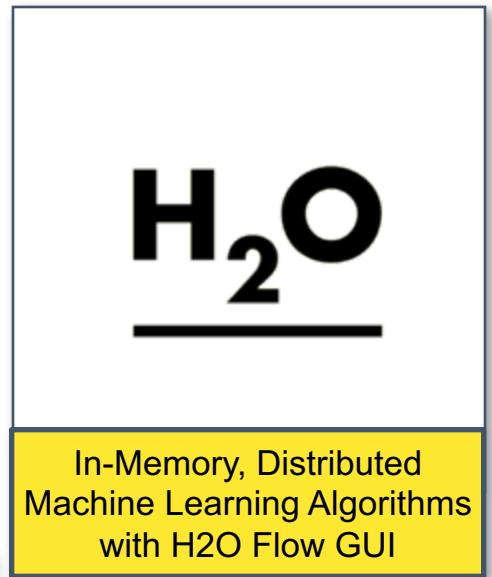
System  
Integrators

Data Stores

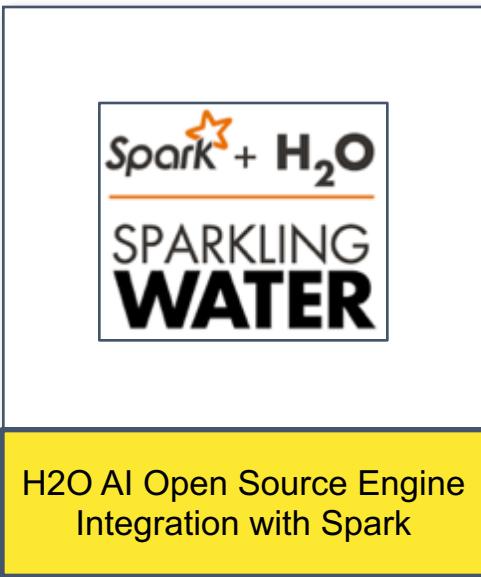


# H2O.ai Product Suite

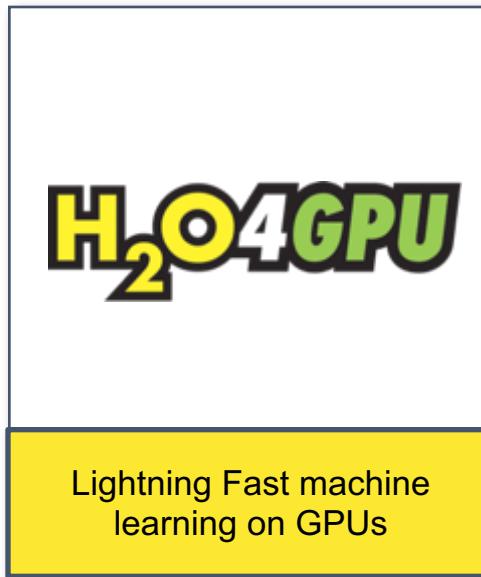
## Open Source



In-Memory, Distributed  
Machine Learning Algorithms  
with H2O Flow GUI

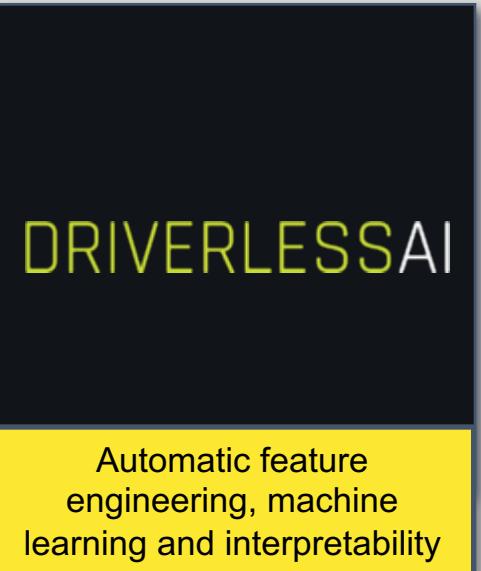


H2O AI Open Source Engine  
Integration with Spark



Lightning Fast machine  
learning on GPUs

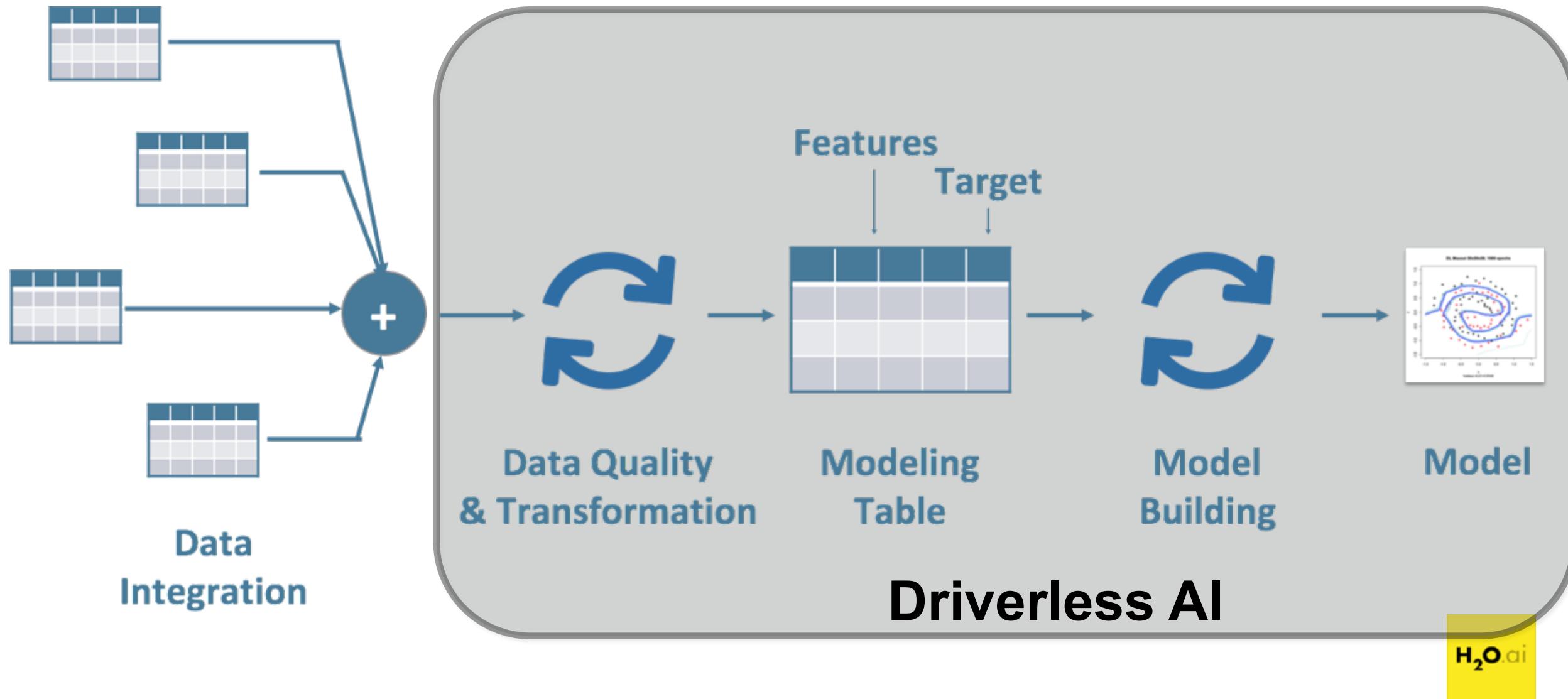
- 100% open source – Apache V2 licensed
- Built for data scientists – interface using R, Python  
on H2O Flow (interactive notebook interface)
- Enterprise Support subscriptions

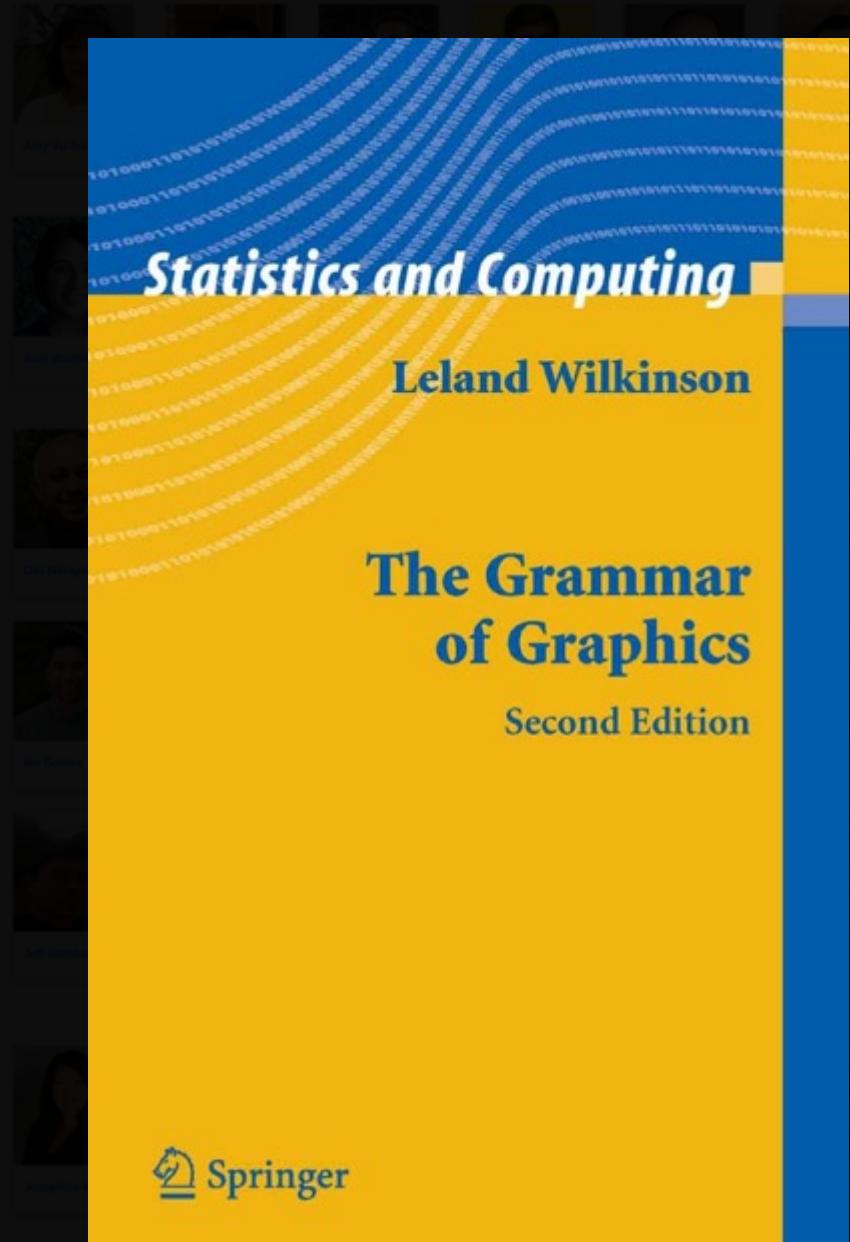


- Enterprise software
- Built for domain users, analysts & data scientists – GUI based interface for end-to-end data science
- Fully automated machine learning from ingest to deployment
- User licenses on a per seat basis (annual subscription)

# Why Driverless AI?

# Driverless AI: Automates Data Science and ML Workflows





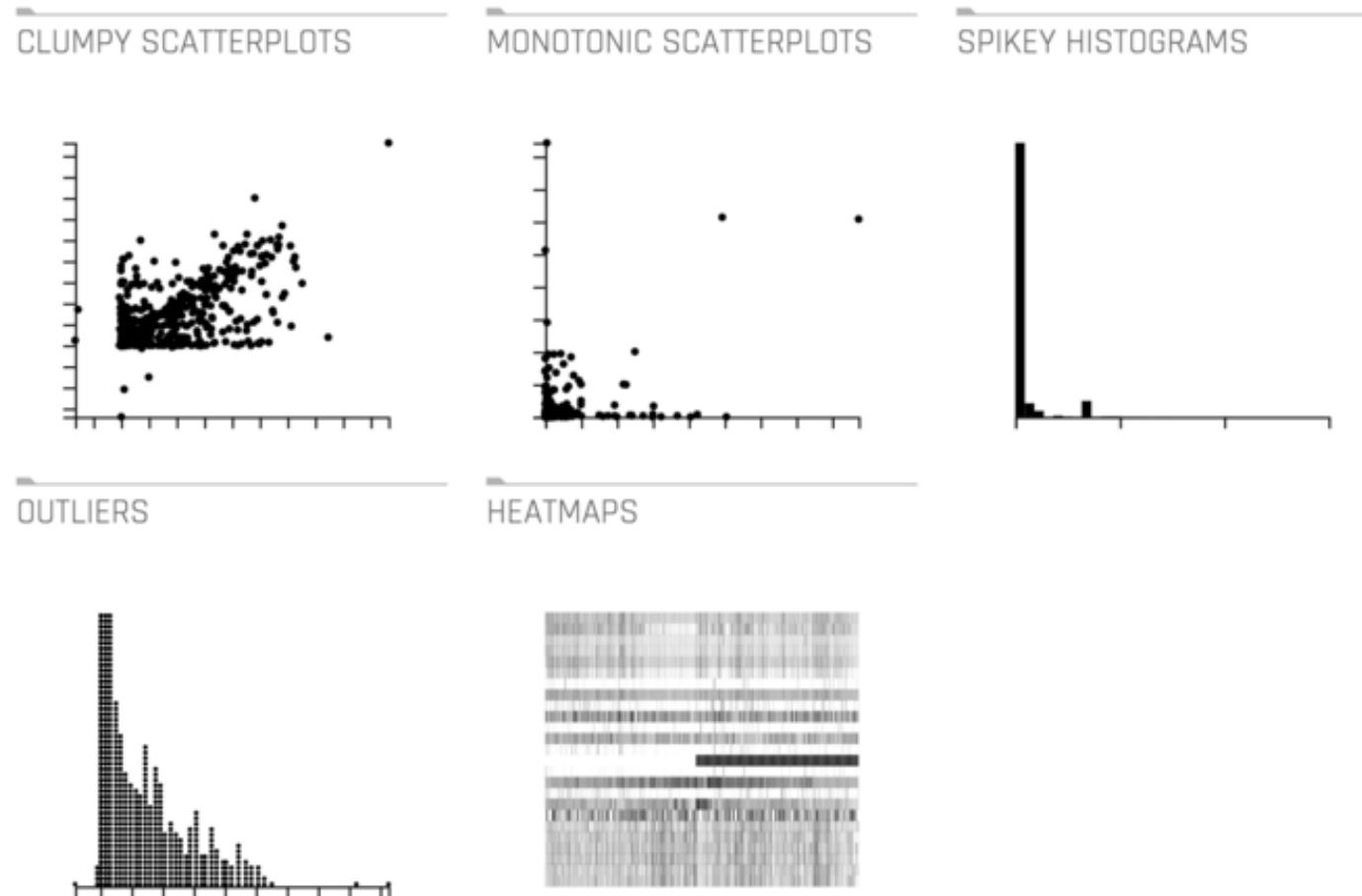
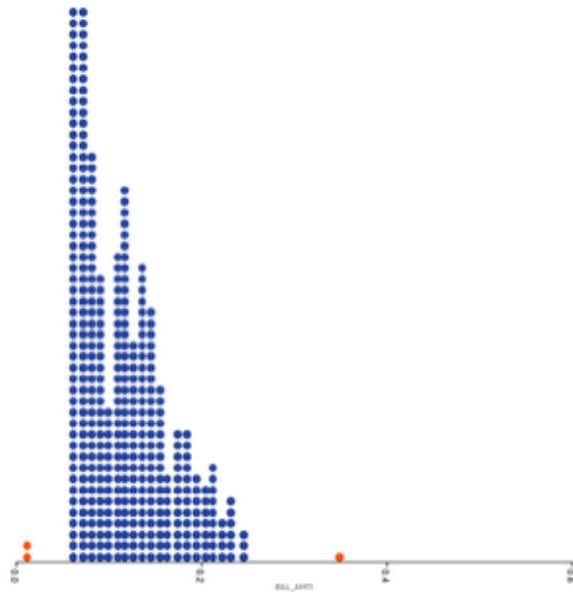
## Origin of R Package `ggplot2`



# Automatic Visualization

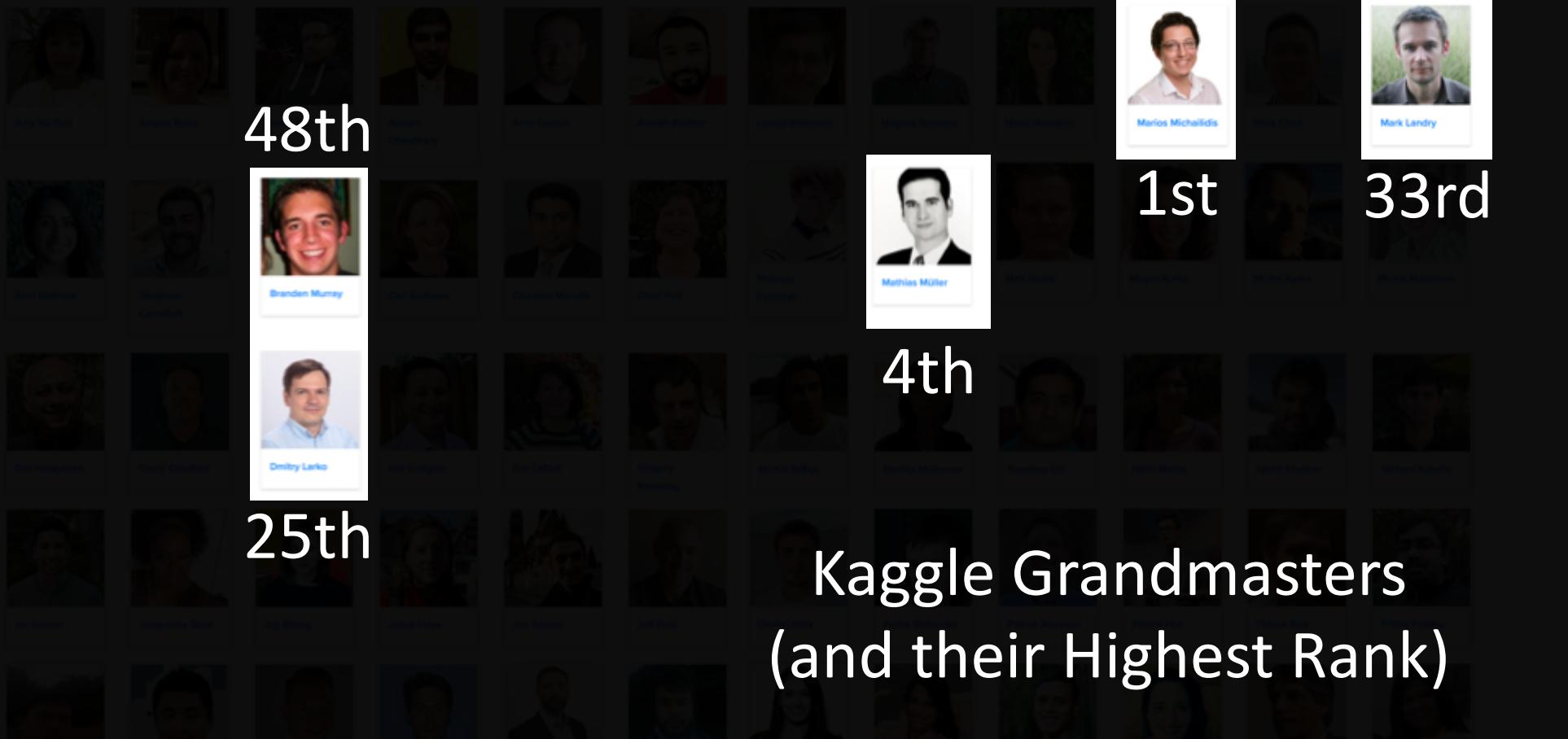
H2O.ai

Automatic Scagnostics and other visualizations to generate the most relevant visualizations for each dataset



"Confidential and property of H2O.ai. All rights reserved"

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



 113  
Grandmasters

 980  
Masters

 3,339  
Experts

 46,135  
Contributors

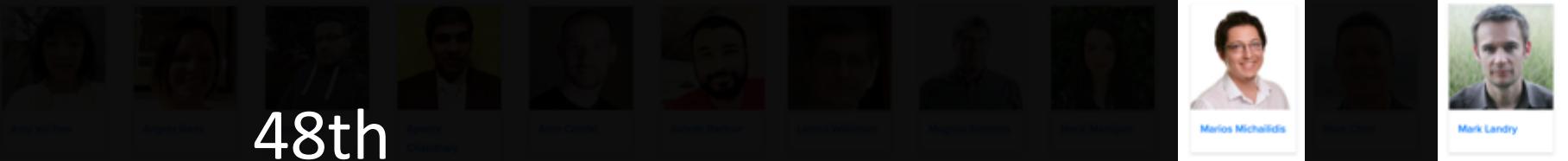
 33,242  
Novices

About 80,000 Kagglers

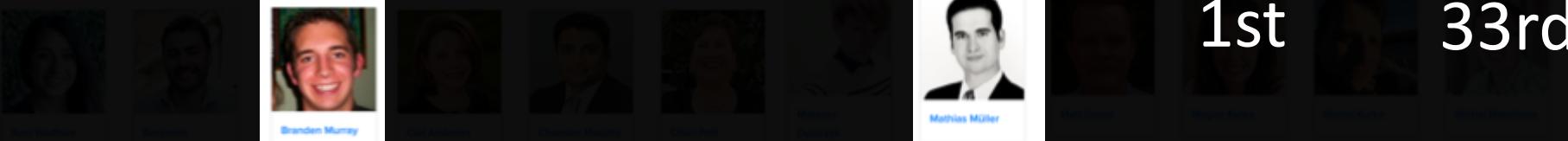
H<sub>2</sub>O Team

13th

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



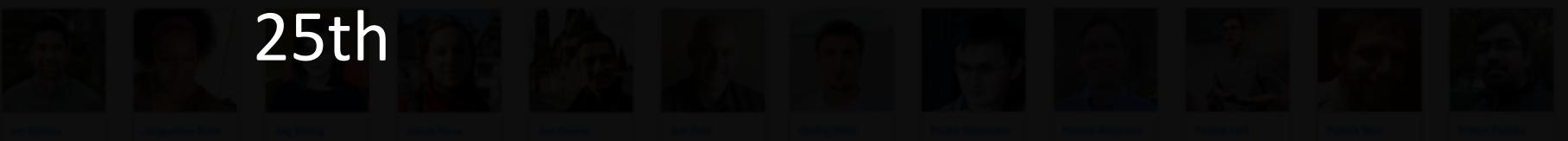
48th



Branden Murray



Dmitry Larko

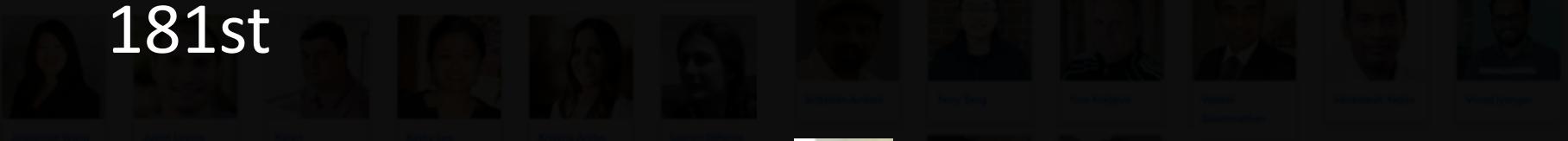


25th



Jo-Fai Chow

Hoping to get closer to them at some point ...



181st



13th

H<sub>2</sub>O Team

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

# Secret Sauce: 1) Grandmaster Feature Engineering

## VARIABLE IMPORTANCE

632_WoE:v56:v79.0	1.00
441_WoE:v13:v66.0	0.25
197_v50	0.20
94_CVTE:v66.0	0.18
556_NumToCatWoE:v50.0	0.13
621_NumToCatWoE:v114:v49:v72:v73.0	0.12
588_InteractionMul:v114:v50	0.11
592_TruncSVD:v12:v50:v85:v90.1	0.09
417_ClusterDist7:v104:v114:v50.1	0.06
535_NumToCatWoE:v114:v25:v39:v49:v95.0	0.05
607_NumToCatTE:v114:v46:v62:v65:v98.0	0.05
616_TruncSVD:v12:v15:v18:v50.1	0.04
438_CVTE:v107:v127:v22:v37:v5.0	0.04
430_ClusterDist7:v49:v50.1	0.04

Numerical/Categorical Interactions, Target Encoding, Clustering, Dimensionality Reduction, Weight of Evidence, etc.

## VARIABLE IMPORTANCE

11_EWMA(0.05)(0)TargetLags:Dept:Store.44:45:51:52:53:105	1.00
18_TargetLog:Dept:Store.52	0.32
22_ClusterDlst4:Dept.3	0.20
13_LagsMax:Dept:Store.Weekly_Sales.44:45:51:52:53:105	0.15
22_ClusterDlst4:Dept.0	0.15
33_Freq:Dept:Store	0.15
18_TargetLog:Dept:Store.44	0.14
22_ClusterDlst4:Dept.1	0.11
18_TargetLog:Dept:Store.45	0.10
2_Freq:Store	0.08
22_ClusterDlst4:Dept.2	0.05
31_EWMA(0.05)(0)TargetLags:Dept.44:45:51:52:53:105	0.05
25_TargetLog:Store.44	0.05
13_LagsMedian:Dept:Store.Weekly_Sales.44:45:51:52:53:1...	0.04

Time-Series: Lags and historical aggregates with causality constraints

# Secret Sauce: 2) Grandmaster Pipeline Tuning + Validation

Example: Driverless AI BNP Paribas on 3-GPU workstation

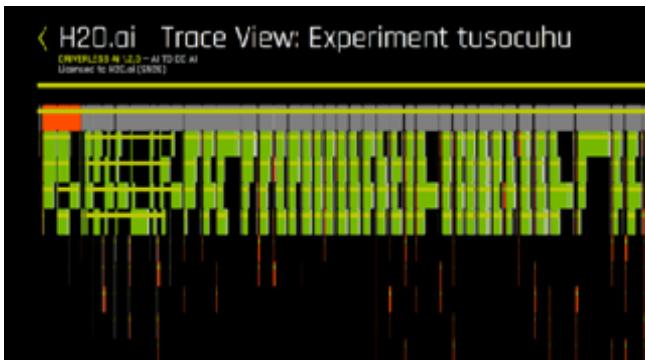
Recipe: AutoDL (171 iterations, 12 individuals)  
Validation scheme: stratified, 1 internal holdout  
Feature engineering: 18923 features tested (344 selected)  
Timing:  
Data preparation: 8.44 secs  
Model parameter tuning: 403.98 secs (19 models trained)  
Feature engineering: 15424.53 secs (1008 models trained)  
Final model training: 1935.21 secs (26 models trained)  
Validation score: LOGLOSS = 0.47811 +/- 0.0023019 (baseline)  
Validation score: LOGLOSS = 0.43681 +/- 0.0037107 (final model)  
Test score: LOGLOSS = N/A (no target)



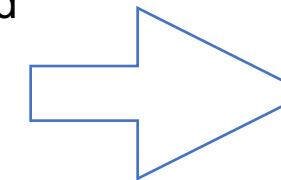
MTV

evolutionary strategies

massively parallel processing  
(multi-CPU, multi-GPU)



19,000 features tested

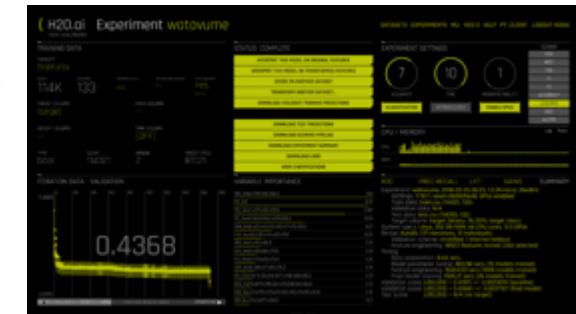
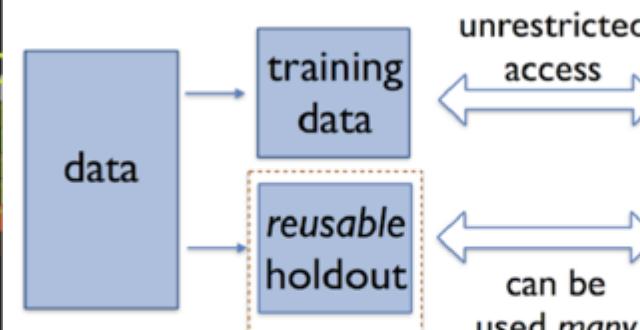


1 final optimal scoring pipeline

1,000 models trained

reliable generalization estimates (overfitting avoidance)

Reusable holdout method



essentially as good as  
using fresh data each time!

# Statistical Learning vs Deep Learning - We Do Both!

**What do these settings mean?**

## ACCURACY

- Training data size: **1,000,000 rows, 8 cols**
  - Feature evolution: **GBM, 5-fold CV**
  - Final pipeline: **Ensemble (5xGBM, 2xTensorFlow), 5-fold CV**

TIME

- Feature evolution: **8 individuals**, up to **47 iterations**
  - Early stopping: After **5** iterations of no improvement

INTERPRETABILITY ■

- Feature pre-pruning strategy: **None**
  - Monotonicity constraints: **disabled**
  - Feature engineering search space (where applicable):  
['Clustering', 'Date', 'FrequencyEncoding', 'Identity',  
'Interactions', 'NumEncoding', 'TargetEncoding', 'Text',  
'TruncatedSVD', 'WeightOfEvidence']

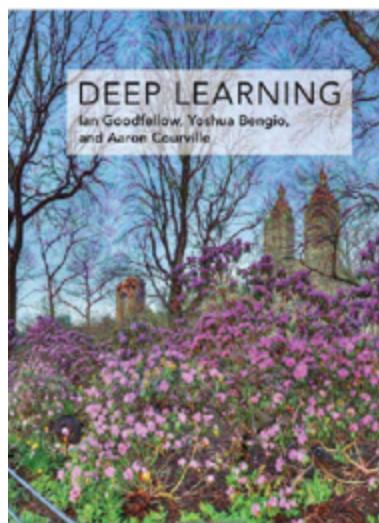
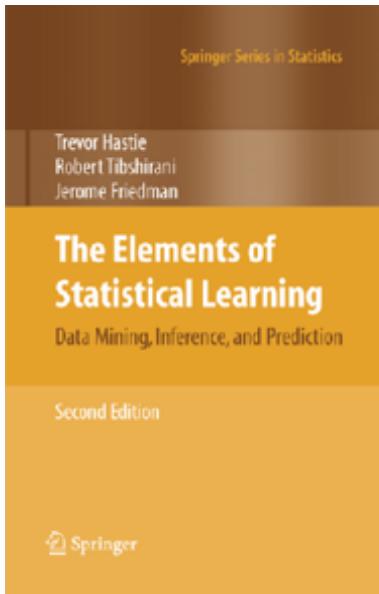
## GBM models to train:

- Model and feature tuning: **165**
  - Feature evolution: **620**
  - Final pipeline: **35**

Estimated max. total memory usage:

- Feature engineering: **2.6GB**
  - GPU XGBoost: **656.0MB**

Estimated runtime: **10 minutes**



## GLM/CART/RF/GBM/XGBoost K-Means/PCA/SVD

Typically better for structured data  
(CSV, SQL, Transactional)

<https://web.stanford.edu/~hastie/Papers/ESLII.pdf>

# TensorFlow Deep Learning

Typically better for unstructured data  
(Images, Video, Audio, Text)

<http://www.deeplearningbook.org>



# Accuracy

- Automatic feature engineering to increase accuracy - AlphaGo for AI
- Automatic Kaggle Grandmaster recipes in a box for solving wide variety of use-cases
- Automatic machine learning to find and tune the right ensemble of models

## Driverless AI: top 5% in Amazon Kaggle competition

Driverless AI produces feature engineering pipeline (“more columns”) for downstream use



Amazon.com - Employee Access

Predict an employee's access needs  
\$5,000 - 1,687 teams - 4 years ago

Driverless AI: 80th place (out of 1687 - top 5%)

## Driverless AI: Top-10 in BNP Paribas Kaggle competition



single run, fully automated: 2h on DGX Station! 6h on PC

BNP Paribas Cardif Claims Management

Can you accelerate BNP Paribas Cardif's claims management process?

\$30,000 - 2,926 teams - 2 years ago

Submission and Description  
sub.csv  
2 minutes ago by Arnaud Candel  
9408bf7/10/1 cv=0.4354 finished after 172 iterations

Private Score: 0.42945  
Public Score: 0.43156

#	In the money	Team Name	Kernel	Team Members	Score (P)	Entries	Last
1	-	Dexter's Lab			0.42037	198	2y
2	-	escalated chi			0.42079	162	2y
3	-	Exploding Kittens			0.42182	524	2y
4	-	Brandon Nickel   utility			0.42259	254	2y
5	-	the flying bunnies brothers			0.42450	264	2y
6	-	n_m			0.42535	4	2y
7	-	PAFY			0.42557	319	2y
8	-	KAIME			0.42688	121	2y
9	-	Jack (Dapper)			0.42744	22	2y
10	+1	Dmitry & Bohdan			0.43000	192	2y
11	+1	Li-Der			0.43006	56	2y
12	+2	BIGM4IPRS			0.43089	338	2y
13	-	x2Red			0.43187	55	2y
14	-	Franchies			0.43146	134	2y
15	+1	Aims			0.43168	55	2y
16	+1	maze numbers			0.43362	164	2y
17	-	BB-B-2			0.43313	129	2y
18	+3	no one			0.43367	68	2y

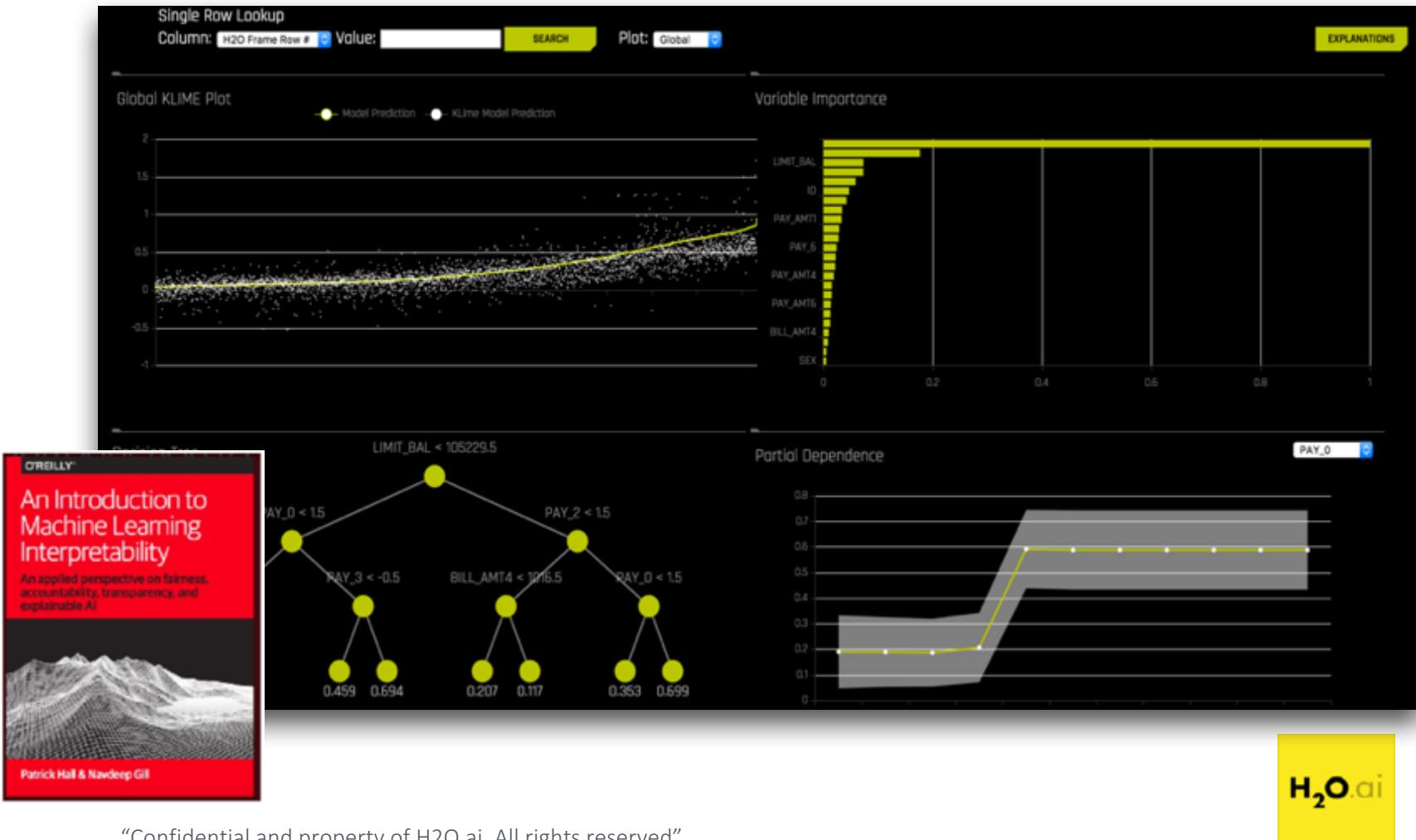
Driverless AI: 10th place in private LB at Kaggle (out of 2926)

2 months for Grandmasters — 2 hours for Driverless AI

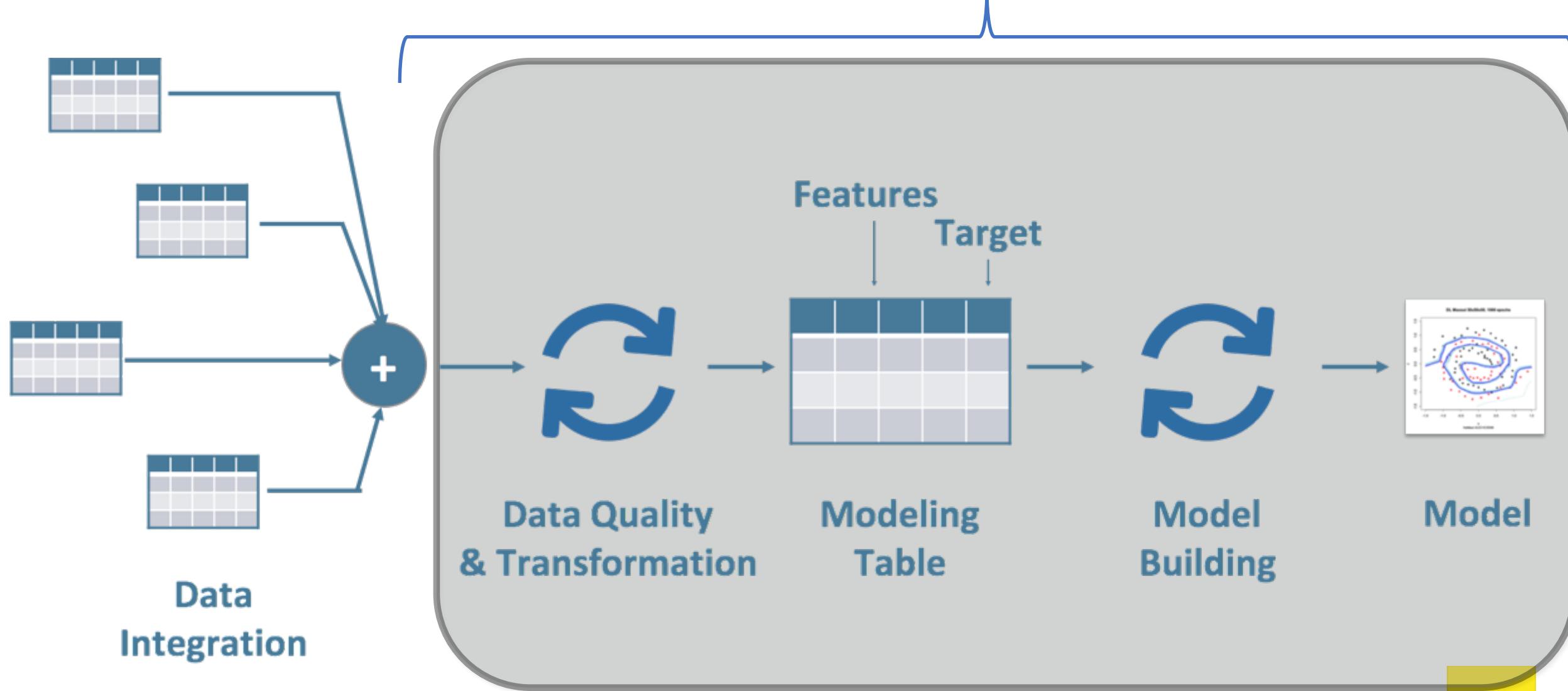


# Interpretability

- Interpretability for debugging, not just for regulators
- Get reason codes and model interpretability in plain english
- K-Lime, LOCO, partial dependence and more



# Deployment: Auto Generated Pipelines



Driverless AI = AI to do AI



## BNP Paribas Cardif Claims Management

Live Demo

Can you accelerate BNP Paribas Cardif's claims management process?

\$30,000 · 2,926 teams · 2 years ago

Overview Data Kernels Discussion Leaderboard Rules Team

My Submissions Late Submission

### Overview

#### Description

As a global specialist in personal insurance, **BNP Paribas Cardif** serves 90 million clients in 36 countries across Europe, Asia and Latin America.

#### Evaluation

In a world shaped by the emergence of new uses and lifestyles, everything is going faster and faster.

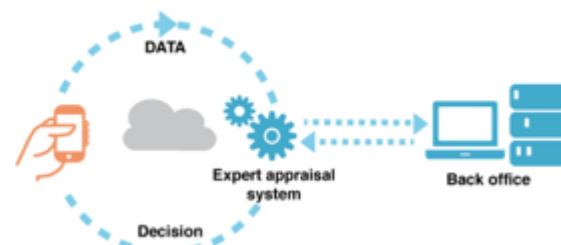
When facing unexpected events, customers expect their insurer to support them as soon as possible.

However, claims management may require different levels of check before a claim can be approved and a payment can be made. With the new practices and behaviors generated by the digital economy, this process needs adaptation thanks to data science to meet the new needs and expectations of customers.

#### Prizes

#### Timeline

#### About Bnp Paribas Cardif



In this challenge, BNP Paribas Cardif is providing an anonymized database with two categories of claims:

1. claims for which approval could be accelerated leading to faster payments
2. claims for which additional information is required before approval

Kagglers are challenged to predict the category of a claim based on features available early in the process, helping BNP Paribas Cardif accelerate its claims process and therefore provide a better service to its customers.

## Binary Classification

	# ID	# target	A v1	A v2	A v3	A v4
1	3	1	1.33573941541	8.72747443554	C	3.9210257481
2	4	1			C	
3	5	1	0.943876910249	5.31007920093	C	4.41096869049
4	6	1	0.797414556191	8.30475713591	C	4.22592985639
5	8	1			C	
6	9	0			C	
7	12	0	0.899805657905	7.31299494722	C	3.49414846822
8	21	1			C	
9	22	0	2.07865125956	8.46261880883		3.73902977371
10	23	1	1.14480237605	5.8806060887	C	3.24446880767
11	24	1			C	
12	27	1			C	
13	28	0			C	
14	30	1	1.40026686648	5.36720439306	C	4.12215483179
15	31	1	2.2600357086	14.6932626703	C	5.15075011304
16	32	1			C	
17	33	1	0.622896057771	7.02473161621	C	4.19368768327

Binary Classification

# Driverless AI Experiment – Live Demo

< H2O.ai Experiment limodida

DRIVERLESS AI 1.3.1 - AI TO DO AI  
Licensed to H2O.ai (SN28)

TRAINING DATA

ASSISTANT

Dataset: BNPParibas-train.csv

Rows: 114K Columns: 133 Dropped Cols: -- Validation Dataset: Yes BNPParibas-test.csv

Target Column: target Fold Column: --

Weight Column: -- Time Column: [OFF]

Type: bool Count: 114321 Unique: 2 Target Freq: 87021

STATUS: COMPLETE

INTERPRET THIS MODEL  
SCORE ON ANOTHER DATASET  
TRANSFORM ANOTHER DATASET...  
DOWNLOAD (HOLDOUT) TRAINING PREDICTIONS

CLASSIFICATION REPRODUCIBLE ENABLE GPUs

EXPERIMENT SETTINGS

SCORER: GINI, MCC, F05, F1, F2, ACCURACY, LOGLOSS, AUC, AUICPR

EXPERT SETTINGS: Accuracy: 5, Time: 1, Interpretability: 5

CPU / MEMORY

CPU: Notifications Log Trace

VARIABLE IMPORTANCE

Variable	Importance
274_NumCatTEv50v66.0	1.00
298_NumCatTEv47v60v66.0	0.79
153_WoEv110v31v56v66v79.0	0.77
322_WoEv110v56v66v79.0	0.67
352_WoEv31v47v56v66v79.0	0.39
398_WoEv110v22v24v31v47v56v66.0	0.22
258_NumCatTEv114v31v50v66.0	0.17
304_Interaction0v114v50	0.16
243_CVTEv110v31v47v56v66.0	0.15
237_WoEv110v24v30v31v47v66.0	0.09
281_WoEv31v47v56v66.0	0.08
254_WoEv22v79.0	0.08
162_NumCatTEv110v39v50.0	0.07
171_Cluster0listBy12v129v50.1	0.05

ROC PRE-RECALL LIFT GAINS SUMMARY

Experiment: limodida, 2018-10-03 22:13, 1.3.1  
Settings: 5/1/5, seed=537069244, GPUs enabled  
Train data: BNPParibas-train.csv (114321, 133)  
Validation data: N/A  
Test data: BNPParibas-test.csv (114393, 132)  
Target column: target (binary, 76.120% target class)  
System specs: Docker/Linux, 480 GB, 32 CPU cores, 8/8 GPUs  
Max memory usage: 9.62 GB, 6.18 GB GPU  
Recipe: AutoDL (3 iterations, 16 individuals)  
Validation scheme: stratified, 1 internal holdout  
Feature engineering: 4492 features tested (463 selected)  
Timing:  
Data preparation: 8.91 secs  
Model and feature tuning: 123.51 secs (17 models trained)  
Feature evolution: 114.86 secs (16 of 24 models trained)  
Final pipeline training: 557.92 secs (4 models trained)  
Validation score: LOGLOSS = 0.46764 +/- 0.0027655 (baseline)  
Validation score: LOGLOSS = 0.45588 +/- 0.0034827 (final pipeline)  
Test score: LOGLOSS = N/A (no target)

© 2017-2018 H2O.ai. All rights reserved.

H2O.ai

# Deployment: Scoring Pipeline Example

```
|>> ls
|          valid license
| README.txt
| example.csv
| license.sig
|>>
|>> bash run_example.sh
=====
| Running MOJO2 example
=====

MOJO file      : pipeline.mojo
Input file     : example.csv
New data (raw features only, no target)

Command line : java -Xmx5g -Dai.h2o.mojos.runtime.license.file= -cp mojo2-runtime.jar ai.h2o.mojos.ExecuteMojo
               pipeline.mojo example.csv

Mojo load time: 17.708 sec
target.0,target.1
0.19523265921932942,0.8047673407806706
0.1944353407855759,0.8055646592144241
0.10768736793959754,0.8923126320604025
0.04318818186408424,0.9568118181359158
0.06411240377340266,0.9358875962265973
0.057372518455691135,0.9426274815443089
0.07560148672603617,0.9243985132739638
0.03610371824346359,0.9638962817565364
0.06139625396388515,0.9386037460361148
0.06348379323463049,0.9365162067653695
Time per row: 3.100 msec (total time: 31.000 msec)
|>>
```

valid license

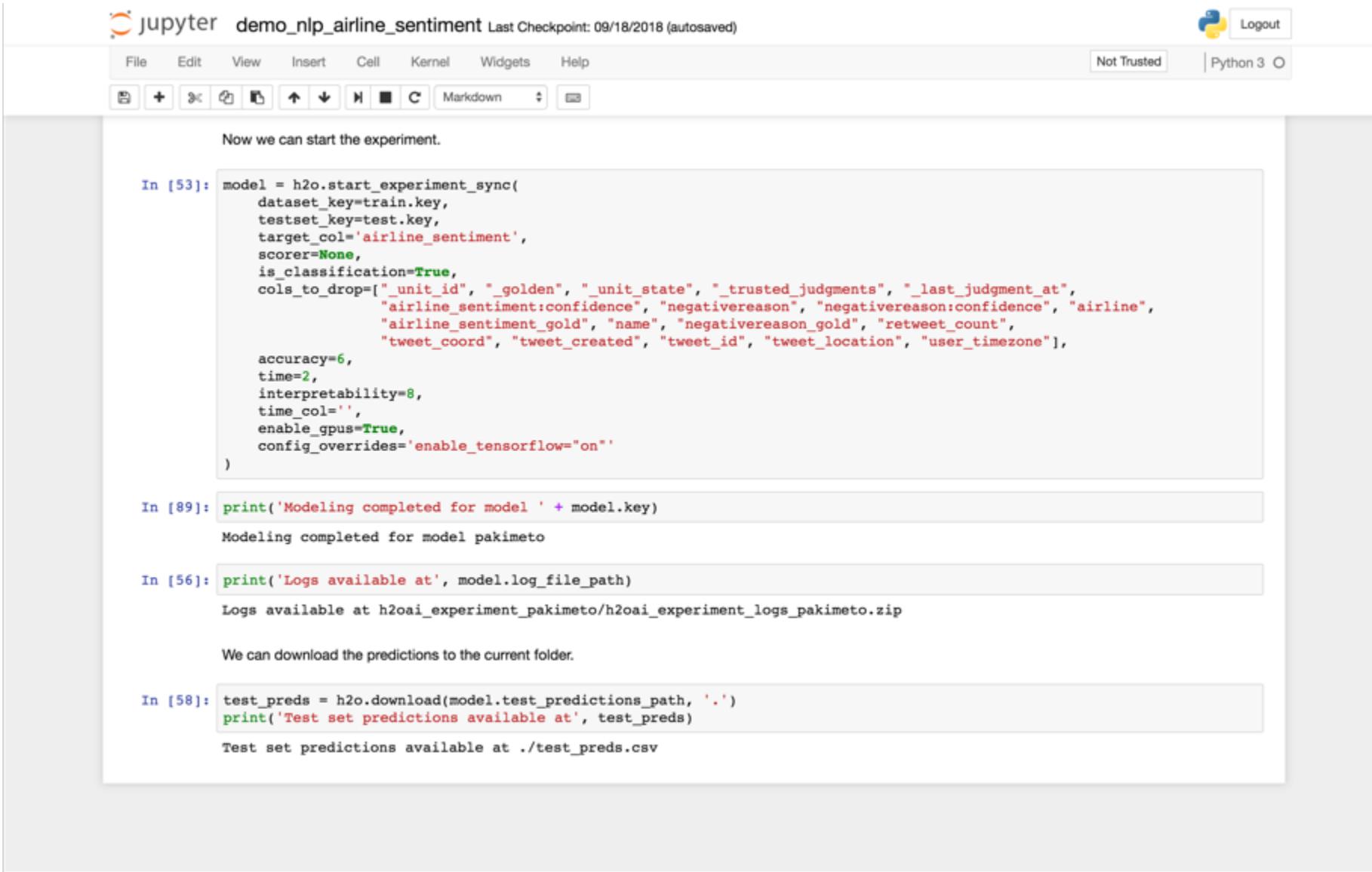
Pipelines generated from Driverless AI experiment

New data (raw features only, no target)

Fast, practical scoring speed in ms  
(including all feature engineering and scoring steps)

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

# Python API: Running Driverless AI with a Script



The screenshot shows a Jupyter Notebook interface with the title "jupyter demo\_nlp\_airline\_sentiment Last Checkpoint: 09/18/2018 (autosaved)". The notebook has a "Not Trusted" status and is using "Python 3". The code cell In [53] contains Python code to start an experiment sync:

```
In [53]: model = h2o.start_experiment_sync(
    dataset_key=train.key,
    testset_key=test.key,
    target_col='airline_sentiment',
    scorer=None,
    is_classification=True,
    cols_to_drop=['_unit_id", "_golden", "_unit_state", "_trusted_judgments", "_last_judgment_at",
                  "airline_sentiment:confidence", "negativereson", "negativereson:confidence", "airline",
                  "airline_sentiment_gold", "name", "negativereson_gold", "retweet_count",
                  "tweet_coord", "tweet_created", "tweet_id", "tweet_location", "user_timezone"],
    accuracy=6,
    time=2,
    interpretability=8,
    time_col='',
    enable_gpus=True,
    config_overrides='enable_tensorflow="on"'
)
```

The code cell In [89] prints the completed model key:

```
In [89]: print('Modeling completed for model ' + model.key)
Modeling completed for model pakimeto
```

The code cell In [56] prints the log file path:

```
In [56]: print('Logs available at', model.log_file_path)
Logs available at h2oai_experiment_pakimeto/h2oai_experiment_logs_pakimeto.zip
```

A note indicates that predictions can be downloaded:

We can download the predictions to the current folder.

The code cell In [58] demonstrates this by downloading the test set predictions:

```
In [58]: test_preds = h2o.download(model.test_predictions_path, '.')
print('Test set predictions available at', test_preds)
Test set predictions available at ./test_preds.csv
```

# Driverless AI on IBM Power

The sidebar on the left is titled "DRIVERLESSAI 1.3.1". It includes a search bar labeled "Search docs". Below the search bar are sections for "RELEASE NOTES", "OVERVIEW", and "INSTALLATION AND UPGRADE". Under "INSTALLATION AND UPGRADE", there is a heading "Installing and Upgrading Driverless AI" followed by a list of options: "Linux Docker Images", "Linux RPMs", "Linux DEBs", "Linux TAR SH", "Linux in the Cloud", "Mac OS X", "Windows 10 Pro", and "IBM Power".

[Docs](#) » [Installing and Upgrading Driverless AI](#) » [IBM Power](#)

[View page source](#)

## IBM Power

Go to <https://www.h2o.ai/download/> to download the IBM Power RPM packages, DEB packages, TAR SH, or Docker image, and follow the steps for the appropriate install type.

- [IBM DEB](#)
- [IBM RPMs](#)
- [IBM TAR SH](#)
- [IBM Docker Images](#)

[Previous](#)

[Next](#)

© Copyright 2017-2018 H2O.ai. Last updated on Sep 25, 2018.

Built with [Sphinx](#) using a [theme](#) provided by [Read the Docs](#).

docs.h2o.ai

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

# Driverless AI on IBM Power

Customer Case Study

## Driving Away Fraudsters at Paypal



- 6% Increase in Model Accuracy*
- 6X Faster Model Development*
- Top 5 Features Created Automatically*



Product Datasheet



## H2O Driverless AI with IBM Power



**H2O Driverless AI on IBM Power** is an automatic machine learning platform that gives you an experienced "data scientist in a box" to create AI driven products and services to transform your business.

store and online sales, and in manufacturing with sensor data to improve supply chain or predictive maintenance.

### Trusted AI Results

Delivering machine learning results your business can trust is a key goal of data science teams. H2O Driverless AI delivers highly accurate models with machine interpretability that helps explain how the models work to the business. Delivering trusted and transparent results increases adoption of AI and also allow your company to comply with government regulations.

### Easy AI Deployment

Model deployment remains one of the most common challenges for data scientists. Models can take weeks or even months to reach production and may be modified to work with production systems. H2O Driverless AI creates ultra-low latency automatic scoring pipelines for easy deployment. In addition, H2O supports training, testing and model versioning so that data science and business teams can work together to bring models from data science to production in minutes, not months.

**H2O Driverless AI on IBM Power**

Scale with  
**2.6X**  
More RAM

Faster Data Ingest  
**2X**  
Max I/O

Accelerate Time-Series  
**5X**  
Faster on IBM Power

# Driverless AI Delivers “Expert Data Scientist in a Box”

- Created and supported by world renowned AI experts
- Empowers companies to accomplish AI and ML with a single platform
- Performs the function of an expert data scientist and adds more power to both novice and expert teams
- Details and highlights insights and interpretability with easy to understand results and visualizations



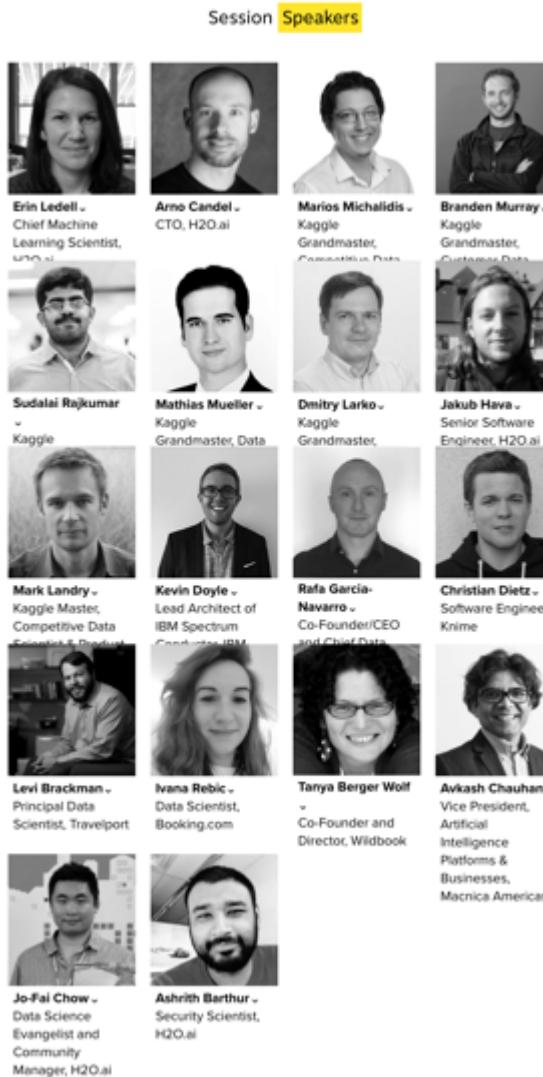
21 day free trial for [Driverless AI](#)

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

# Our Flagship Community Event – H2O AI World is finally coming to London!



29<sup>th</sup> & 30<sup>th</sup> Oct, London



More real-world use cases + All H<sub>2</sub>O Kaggle Grandmasters + Hands-on Training

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

# Thanks!



- More Info, Code, and Slides
  - [bit.ly/  
h2o\\_meetups](https://bit.ly/h2o_meetups)
- Contact
  - [joe@h2o.ai](mailto:joe@h2o.ai)
  - [@matlabulous](https://twitter.com/matlabulous)
  - [github.com/woobe](https://github.com/woobe)