

Introduction to H₂O Driverless AI

Use Cases and Live Demos

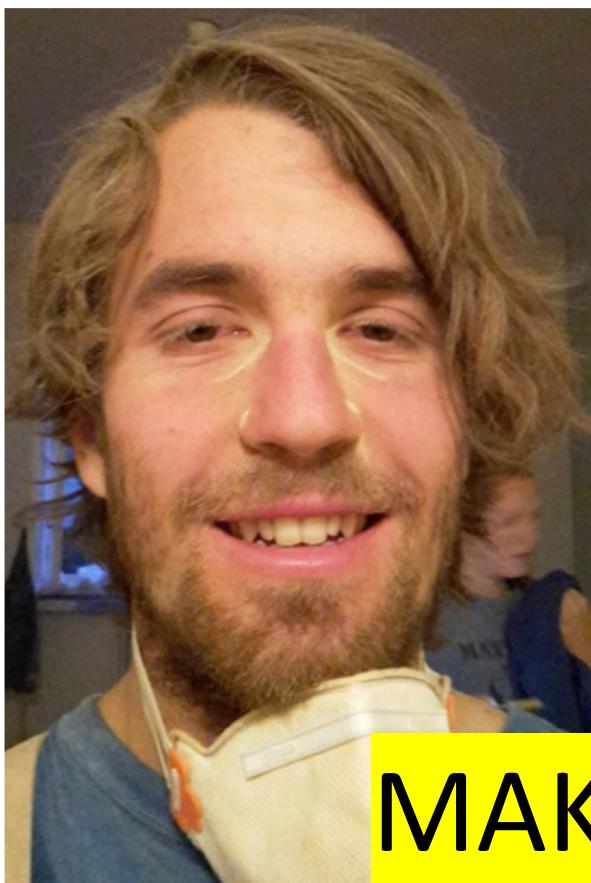


Stefan Pacinda
Software / Sales Engineer

stefan@h2o.ai

More Info → [https://bit.ly/
h2o_meetups](https://bit.ly/h2o_meetups)

About Me



... The harsh reality of startup life ...

- **Before H₂O**

- Writing optimization software at a startup
- (left too early to become a billionaire ☹)
- Writing service virtualization software at a corporation
- (left right on time to catch H2O's wave to ride ☺)

- **At H₂O ...**

- Software Engineer
- Sales Engineer / Solution Architect
- Aspiring Speaker

Reminder: #360Selfie

Partner Ecosystem

SCAN
computers



PENGUIN
COMPUTING

AMAX

SUPERMICR

amazon
web services

Google
Cloud Platform

Microsoft Azure

ORACLE
CLOUD

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

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

Why Driverless AI?

H₂O.ai

CONFIDENTIAL

Driverless AI Delivers **AI for Enterprise**

Talent

Top 10

Data Science Experts

Kaggle Grandmasters

Time

Months
to Hours

GPU accelerated ML
Automatic Pipelines

Time to Insight

Trust

MLI

Auto Doc

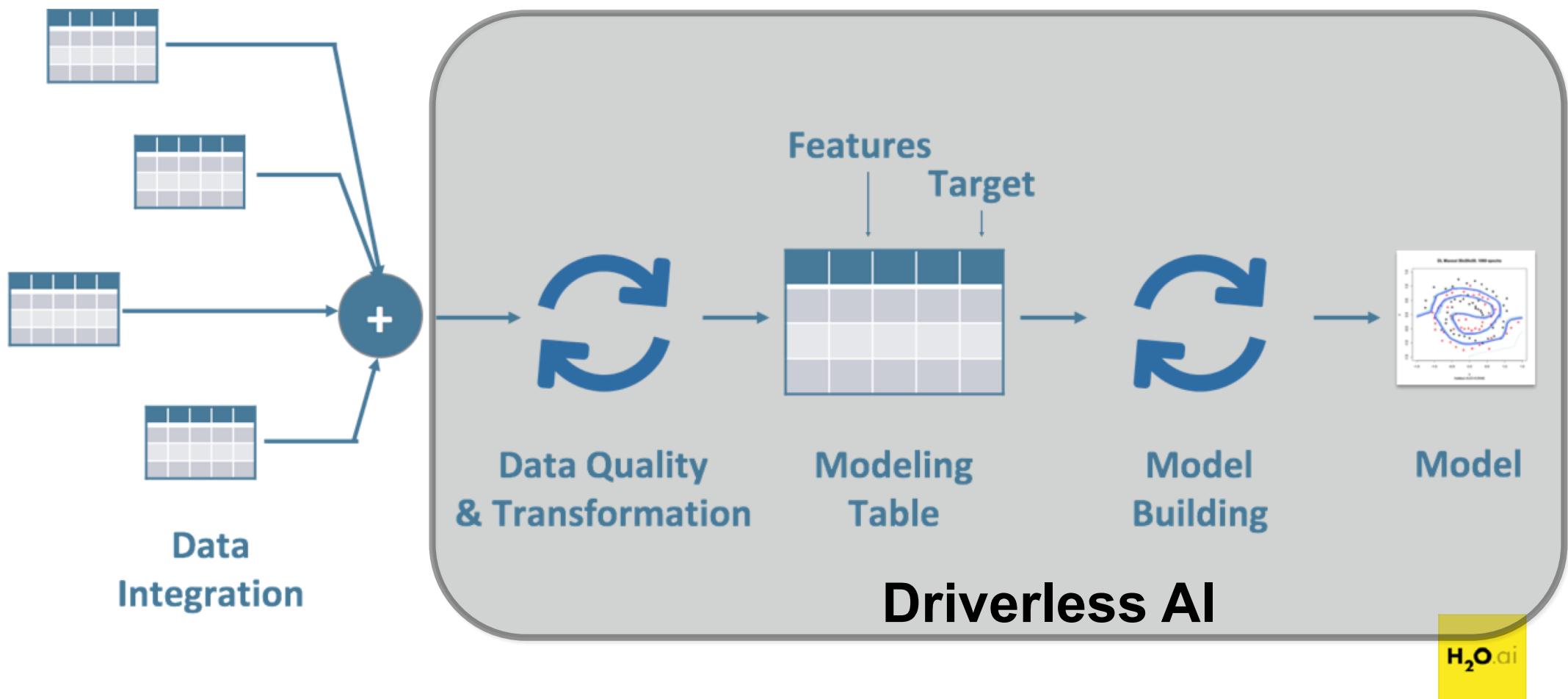
Auto Visualization

Explainability & Transparency

H₂O.ai

CONFIDENTIAL

Driverless AI: Automates Data Science and ML Workflows

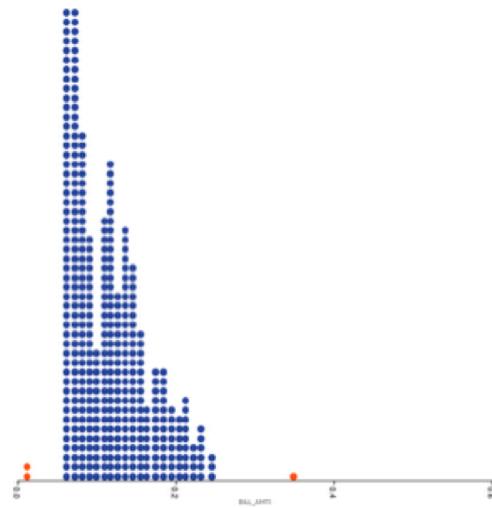


CONFIDENTIAL

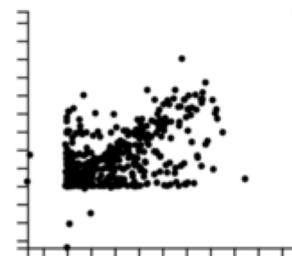
Automatic Visualization

H2O.ai

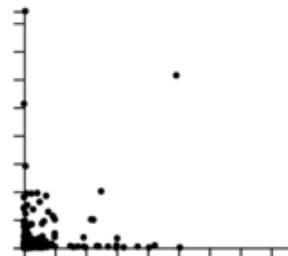
Automatic Scagnostics and other visualizations to generate the most relevant visualizations for each dataset



CLUMPY SCATTERPLOTS



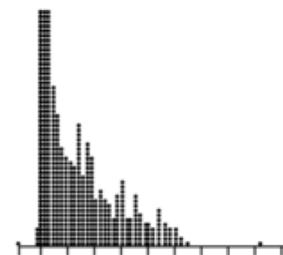
MONOTONIC SCATTERPLOTS



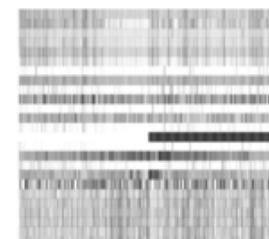
SPIKEY HISTOGRAMS



OUTLIERS



HEATMAPS



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

H₂O.ai

Secret Sauce: 1) Grandmaster Feature Engineering

VARIABLE IMPORTANCE

632_WoE:v56:v79.0	1.00
441_WoE:v113:v66.0	0.25
197_y50	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_TargetLag:Dept:Store.52	0.32
22_ClusterDist4:Dept.3	0.20
13_LagsMax:Dept:Store.Weekly_Sales.44:45:51:52:53:105	0.15
22_ClusterDist4:Dept.0	0.15
33_Freq:Dept:Store	0.15
18_TargetLag:Dept:Store.44	0.14
22_ClusterDist4:Dept.1	0.11
18_TargetLag:Dept:Store.45	0.10
2_Freq:Store	0.08
22_ClusterDist4: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

CONFIDENTIAL

H₂O.ai

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 Attrition

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	Private Score	Public Score
sub.csv 2 minutes ago by Arno Candel 94009f73ff1cv 0.4354 finished after 172 itera	0.42945	0.43156

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

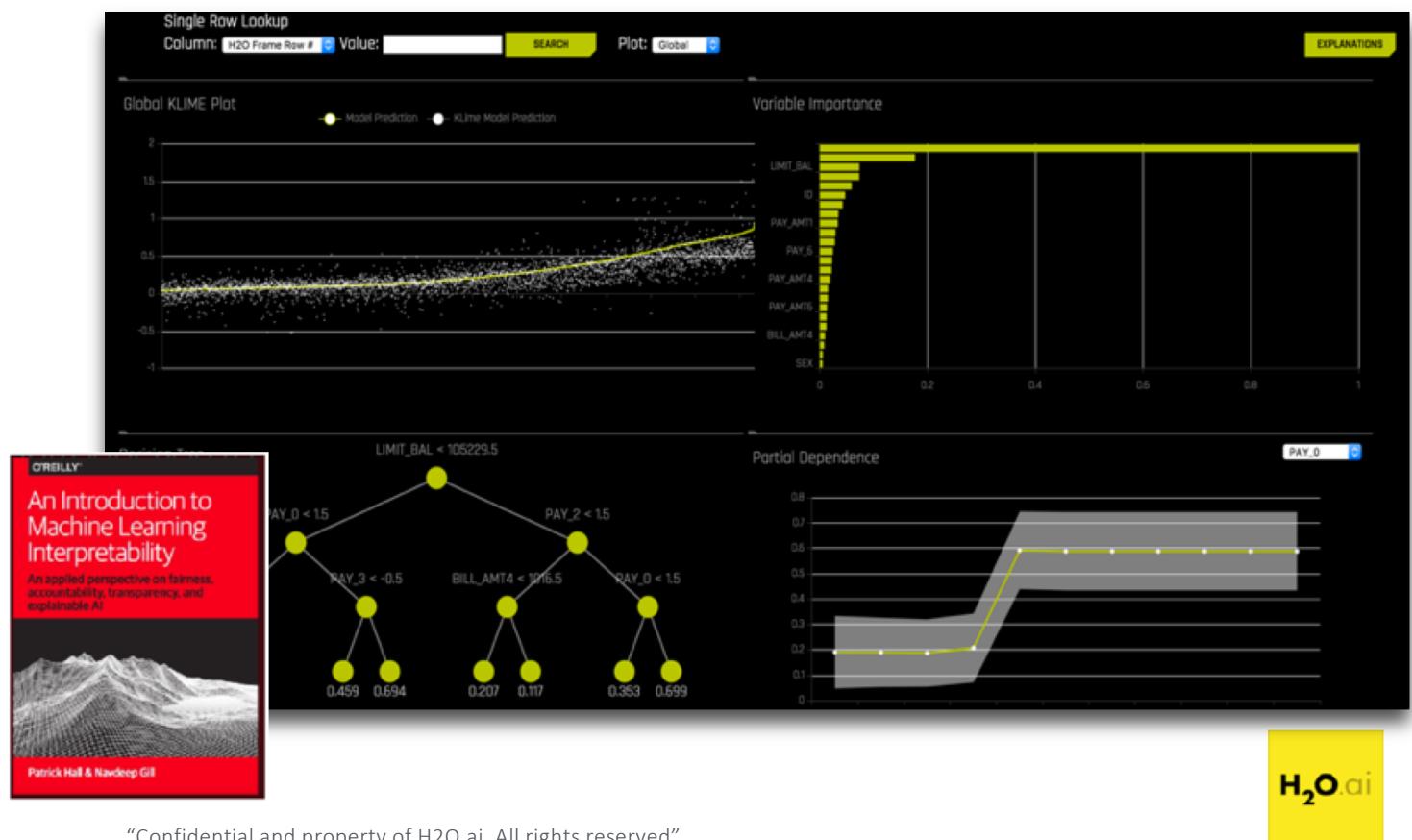
2 months for Grandmasters — 2 hours for Driverless AI



H2O.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



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

Live Demo

H₂O.ai

CONFIDENTIAL

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](#)

H₂O.ai

CONFIDENTIAL