Accelerating AI Deployment with H₂O 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 | • H2O Open Source Machine Learning (14,000 organizations)  
• H2O Driverless AI – Automatic Machine Learning |
| 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
Growing Worldwide Open Source Community

14,000 Companies using H₂O

155,000 data scientists

222 of the 500 Fortune Companies use H₂O

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

Get the Gartner Magic Quadrant here
Partner Ecosystem

HW Vendors  Cloud Providers  Strategic Partners  Value Added Resellers  System Integrators  Data Stores

“Confidential and property of H2O.ai. All rights reserved”
H2O.ai Product Suite

Open Source

- H2O: In-Memory, Distributed Machine Learning Algorithms with H2O Flow GUI
- Spark + H2O: H2O AI Open Source Engine Integration with Spark
- H2O4GPU: 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

Automatic feature engineering, machine learning and interpretability

- 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

Data Integration → Data Quality & Transformation → Modeling Table → Model Building → Model

Features → Target

Driverless AI
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”
Kaggle Grandmasters (and their Highest Rank)

About 80,000 Kagglers

H₂O Team
Hoping to get closer to them at some point ...
Secret Sauce: 1) Grandmaster Feature Engineering

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

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)

19,000 features tested
1,000 models trained
reliable generalization estimates (overfitting avoidance)

Reusable holdout method

DOI: 10.1126/science.aaa9375
Statistical Learning vs Deep Learning - We Do Both!

- **GLM/CART/RF/GBM/XGBoost**
  - Typically better for structured data (CSV, SQL, Transactional)

- **K-Means/PCA/SVD**

- **TensorFlow Deep Learning**
  - Typically better for unstructured data (Images, Video, Audio, Text)
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
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”
Deployment: Auto Generated Pipelines

Driverless AI = AI to do AI
BNP Paribas Cardif Claims Management

Can you accelerate BNP Paribas Cardif’s claims management process?
$30,000 · 2,926 teams · 2 years ago

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.

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.

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.

https://www.kaggle.com/c/bnp-paribas-cardif-claims-management
## Binary Classification

<table>
<thead>
<tr>
<th>#</th>
<th>ID</th>
<th>target</th>
<th>v1</th>
<th>v2</th>
<th>v3</th>
<th>v4</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>3</td>
<td>1</td>
<td>1.3573941541</td>
<td>8.72747443554</td>
<td>C</td>
<td>3.9210257481</td>
</tr>
<tr>
<td>2</td>
<td>4</td>
<td>1</td>
<td>0.943876910249</td>
<td>5.31007928093</td>
<td>C</td>
<td>4.41096869049</td>
</tr>
<tr>
<td>3</td>
<td>5</td>
<td>1</td>
<td>0.797414556191</td>
<td>8.30475713591</td>
<td>C</td>
<td>4.22592985639</td>
</tr>
<tr>
<td>4</td>
<td>6</td>
<td>1</td>
<td>0.899885657985</td>
<td>7.31299494722</td>
<td>C</td>
<td>3.49414865822</td>
</tr>
<tr>
<td>5</td>
<td>8</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>6</td>
<td>9</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>7</td>
<td>12</td>
<td>0</td>
<td>2.07865125956</td>
<td>8.46261880883</td>
<td>C</td>
<td>3.73902977371</td>
</tr>
<tr>
<td>8</td>
<td>21</td>
<td>1</td>
<td>1.14480237605</td>
<td>5.888606888</td>
<td>C</td>
<td>3.24466880767</td>
</tr>
<tr>
<td>9</td>
<td>22</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>10</td>
<td>23</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>11</td>
<td>24</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>12</td>
<td>27</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>13</td>
<td>28</td>
<td>0</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>14</td>
<td>30</td>
<td>1</td>
<td>1.40026686648</td>
<td>5.36720439306</td>
<td>C</td>
<td>4.12215483179</td>
</tr>
<tr>
<td>15</td>
<td>31</td>
<td>1</td>
<td>2.2600357886</td>
<td>14.6932626703</td>
<td>C</td>
<td>5.15875011304</td>
</tr>
<tr>
<td>16</td>
<td>32</td>
<td>1</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>17</td>
<td>33</td>
<td>1</td>
<td>0.622896057771</td>
<td>7.02473161621</td>
<td>C</td>
<td>4.19368768327</td>
</tr>
</tbody>
</table>
Driverless AI Experiment – Live Demo
Deployment: Scoring Pipeline Example

>> ls
README.txt  example.csv  license.sig
>>
>> bash run_example.sh
Running MOJO2 example

MOJO file    : pipeline mojo
Input file   : example.csv

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
0.1944353407855759,0.80556465921144241
0.1976873679399754,0.8923126320604025
0.04318818186408424,0.9568118181359158
0.0641124037340266,0.9358875962265973
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)

Pipelines generated from Driverless AI experiment

Valid license

New data (raw features only, no target)

Fast, practical scoring speed in ms (including all feature engineering and scoring steps)
Python API: Running Driverless AI with a Script

```
In [53]: model = h2o.start_experiment_sync(
    database_key=train.key,
    testset_key=test.key,
    target_col='airline_sentiment',
    accuracy=1,
    time=1,
    interpretability=8,
    time_stamp='',
    enable_opus=True,
    config_overrides='enable_tensorflow=True'
)

In [59]: print('Modeling completed for model ' + model.key)
Modeling completed for model tkn

In [66]: print('Logs available at', model.log_file_path)
Logs available at h2oai_experiment_pakimoto/h2oai_experiment_logs_pakimoto.zip

We can download the predictions to the current folder.

In [68]: test_preds = h2o.download(model.test_predictions_path, '.csv')
print('Test set predictions available at', test_preds)
Test set predictions available at ./test_preds.csv
```
Driverless AI on IBM Power

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

Docs » Installing and Upgrading Driverless AI » IBM Power


Built with Sphinx using a theme provided by Read the Docs.

docs.h2o.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

Delivering AI at Scale
Businesses everywhere have realized that their unique data is key to competitive success and now want to put that data to work with AI. To scale, data science teams need to adopt new tools and techniques that will allow them to get better results and quickly deliver more insights to the business.

Faster AI Development
Increasing the business impact of AI by solving a wider variety of business problems is a key goal of every successful data science team. H2O Driverless AI is optimized to run with GPU acceleration and automates key portions of the data science process including feature engineering and parameter tuning to dramatically reduce the time needed to produce accurate models.

Time Series Helps Forecast Sales, Predict Industrial Machine Failure and More
With the time series capability in Driverless AI, H2O AI directly addresses some of the most pressing concerns of organizations across industries for use cases such as transactional data in capital markets, retail to track in-store and online sales, and in manufacturing with sensor data to improve supply chain or predictive maintenance.

Trusted AI Results
Delivering machine learning results you 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 allows 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 scaling 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

<table>
<thead>
<tr>
<th>Feature</th>
<th>IBM Power Improvement</th>
</tr>
</thead>
<tbody>
<tr>
<td>Scale with</td>
<td>2.6X</td>
</tr>
<tr>
<td>Faster Data Ingestion</td>
<td>2X</td>
</tr>
<tr>
<td>Accelerate Time Series</td>
<td>5X</td>
</tr>
</tbody>
</table>

© 2019 H2O.ai all rights reserved
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
Our Flagship Community Event – H2O AI World is finally coming to London!

29th & 30th Oct, London

More real-world use cases + All H2O Kaggle Grandmasters + Hands-on Training
Thanks!

• More Info, Code, and Slides
  • bit.ly/h2o_meetups

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