

H2O Driverless AI



H2O Driverless AI is an automatic machine learning platform that uses “AI to do AI” to empower data science teams to scale and implement their AI strategy.

Challenges of AI Adoption: Talent, Time and Trust

As organizations look to streamline decision making and improve customer experiences with AI, they are running into three core challenges: talent, time, and trust. First, there is not enough data science talent to build every use case by hand. Even with the right people, hand-coding takes too much time, is not repeatable, and can be error-prone. Each model then has to be explained and validated by the business so users can trust the decisions the model supports. The key to breaking through the talent, time, and trust barriers is advanced machine learning automation with H2O Driverless AI.

Filling the Talent Gap

Data scientists are in short supply for all but the largest technology companies. With Driverless AI, expert and novice data scientists can develop highly accurate models that are ready to deploy. H2O Driverless AI is an award-winning automatic machine learning (AutoML) platform that embeds best practices from the world's leading data scientists into every model. Driverless AI uses a unique evolutionary competition that finds the best combination of features, algorithms, and tuning parameters for each use case. Built-in best practices and guardrails ensure that models don't overfit the data, and prevent other common issues that novice data scientists overlook. Driverless AI allows companies to tackle more use cases with the talent they have or can easily find.

More Models in Less Time

Reducing the time that it takes to develop accurate, production-ready models is critical to delivering AI at scale. Driverless AI automates time-consuming data science tasks including, advanced feature engineering, model selection, hyperparameter tuning, model stacking, and creates an easy to deploy, low latency scoring pipeline. With high-performance computing using both CPUs and GPUs, Driverless AI compares thousands of combinations and iterations to find the best model in minutes or hours. Even advanced data scientists can use Driverless AI to explore more techniques, feature combinations, and tuning parameters than they would be able to do on their own. Driverless AI also streamlines model deployment with automatic scoring pipelines that include everything needed to run the model in production, taking the process from experimentation to production from months to days.

Trusted AI Results

For organizations to adopt AI at scale, data teams, business leaders, and regulators must be able to explain, interpret, and trust AI results. H2O Driverless AI delivers industry-leading capabilities for understanding, debugging, and sharing model results, including Machine Learning Interpretability (MLI) and fairness dashboards, automated model documentation, and reason codes for each prediction for service representatives and customers. With Driverless AI, data teams have everything they need to build trust with business stakeholders and regulators.

Key Capabilities of H2O Driverless AI

Exploratory Data Analysis (AutoViz)

Based on The Grammar of Graphics, the automatic visualizations in Driverless AI provide robust EDA capabilities by automatically selecting data plots based on the most relevant data statistics based on the data shape. Advanced users can also customize visualizations to meet their needs. AutoViz helps users discover trends and issues such as large numbers of missing values or significant outliers that could impact modeling results.

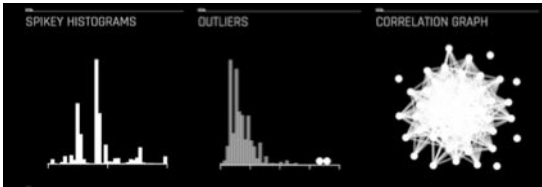


Figure 1: Sample Autoviz Charts Selected Based on Data Shape

Automatic Feature Engineering and Model Building

Feature engineering is the secret weapon that advanced data scientists use to extract the most accurate results from algorithms. H2O Driverless AI uses a unique evolutionary approach to automatically find new, high-value features and feature combinations for a given data set that would be virtually impossible to find using manual methods. Included in the interface is an easy to read variable importance chart that shows the significance of original and newly engineered features.

Machine Learning and Deep Learning

Driverless AI includes leading open-source transformers, embeddings, and frameworks for machine learning and deep learning techniques to handle various data science use cases. With Driverless AI, users can automatically build models for iid data, images, text, and more. For example, Driverless AI includes TensorFlow CNNs for image modeling and NLP libraries from PyTorch, including BERT and other state-of-the-art techniques.

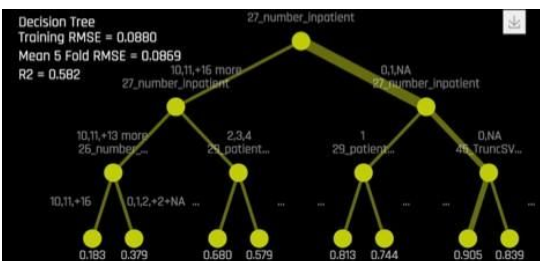


Figure 2: MLI Charts Example, Decision Tree Surrogate Model

Machine Learning Interpretability (MLI)

Driverless AI provides robust explainability and fairness analysis for machine learning models including, K-LIME, LIME-SUP, Shapley on original and engineered features, Variable Importance, Decision Tree Surrogate, ICE, and Partial Dependence Plots. Each of these techniques helps to explore and demystify modeling results.

Driverless AI now also includes straightforward disparate impact analysis to test for model bias. Driverless AI also provides reason codes for every prediction. Maximum transparency and minimal disparate impact are crucial differentiators for those who must justify their models to business stakeholders and regulators.

Automatic Model Documentation (Auto Report)

Data scientists must document the data, algorithms, and processes used to create machine learning models for business users and regulators. Driverless AI automatic model documentation relieves the user from the time-consuming task of recording and summarizing their workflow while building machine learning models. The documentation includes details about the data used, the validation schema selected, model and feature tuning, MLI, and the final model created. AutoDoc saves data scientists time and removes tedious work so they can spend more time doing data science and drive more value for the business.

Automatic Scoring Pipelines

H2O Driverless AI automatically generates both Python scoring and Java low latency scoring pipelines. The scoring pipeline is a unique technology that deploys the feature engineering and the winning machine learning model or ensemble in a highly optimized format that can be deployed anywhere. This technology is critical for enterprises running models that need fast scoring for real-time applications running on a range of devices.

Bring-Your-Own Recipes to Make Your Own AI

Advanced data scientists now can easily extend Driverless AI with customizations that run within the Driverless AI platform, including data preparation, models, transformers, and scorers. These customizations, called recipes, are then treated as first-class citizens in the automatic machine learning optimization process and eventually creating the winning model. Data science teams can explore and use open-source recipes from H2O.ai and other organizations to improve models. They can also develop customizations specific to their use-cases, industry, or their business.

Time Series Recipes

Time-series forecasting is one of the biggest challenges in data science. Time-series models address critical use cases, including demand forecasting, infrastructure monitoring, and predictive maintenance based on the transaction, log, and sensor data. H2O Driverless AI delivers superior time series capabilities to optimize for almost any prediction time window and includes various techniques and models, including world-class recipes for forecasting and epidemic response.

Enterprise-Ready

H2O Driverless AI is scalable, secure, and runs in the cloud or on-premise. The H2O.ai team includes some of the world's leading data scientists and experts in machine learning at scale, and Driverless AI customers enjoy a full range of support, training, and expertise to assist them with their AI journey.