

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.

Delivering AI at Scale

Businesses are creating and driving an AI strategy to gain a competitive edge. There are three critical challenges to achieving AI at scale, including addressing the talent gap, the time a model takes to train, tune and deploy, and being able to trust the AI. H2O Driverless AI is an award-winning and industry-leading automatic machine learning platform that empowers data science teams to scale and deliver trusted, production-ready models to fulfill every business AI strategy and initiatives.

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, data engineers, domain scientists, mathematicians, and statisticians in all businesses can develop highly accurate models that are ready to deploy. H2O Driverless AI was created and developed by expert data scientists so that they can use the platform to perform the tasks of an advanced data scientist automatically. This enables your data scientist to focus on evaluating results and exploring new use cases.

More Models in Less Time

Reducing the time that it takes to develop accurate, production-ready models is critical to solving a large number of business challenges with AI. Driverless AI automates time-consuming data science tasks including, advanced feature engineering, model selection, hyperparameter tuning, model stacking, and model deployment. These processes can be accelerated with high-performance computing on GPU and CPU systems that allow for thousands of combinations and iterations to be tested and to find the best model in hours, not months. Model deployment also is streamlined with automatic scoring pipelines that include everything that is needed to run the model in production.

Trusted AI Results

To adopt AI models at scale, business teams and regulators must be able to interpret and trust AI results. H2O Driverless AI delivers highly accurate models, but also provides vital capabilities for understanding, debugging and sharing model results including Machine Learning Interpretability (MLI) dashboards, automated model documentation and reason codes for service representatives and customers.

Key Capabilities of H2O Driverless AI

AutoVis - Exploratory Data Analysis

H2O Driverless AI automatically selects data plots based on the most relevant data statistics to help users understand their data prior to the model building process. This is useful for understanding the composition of very large data sets and discovering trends and possible issues such as large numbers of missing values or significant outliers that could impact modeling results. It also provides recommendations for transformations to address the problems identified.

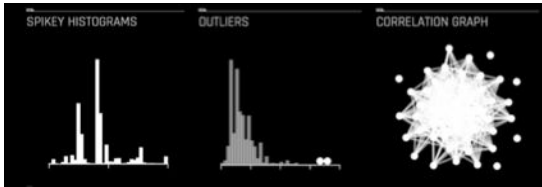


Figure 1: Sample Autoviz Charts Selected Based on Data Shape

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 employs a library of algorithms and feature transformations to automatically engineer new, high-value features for a given data set. Included in the interface is an easy to read variable importance chart that shows the significance of original and newly engineered features.

Machine Learning Interpretability (MLI)

Driverless AI provides robust explainability for machine learning models including, K-LIME, LIME-SUP, Shapley, 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 algorithmic discrimination. Maximum transparency and minimal disparate impact are crucial differentiators for those who must explain and justify their models to business stakeholders and regulators.

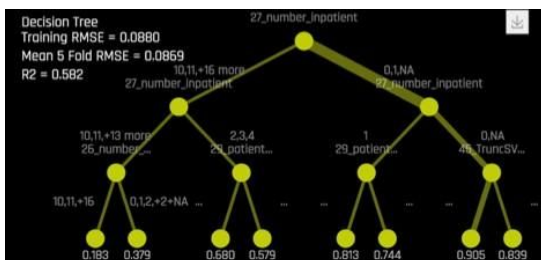


Figure 2: MLI Charts Example, Decision Tree Surrogate Model

Automatic Reason Codes

For many regulated industries, an explanation is required for significant decisions relating to customers, e.g., credit denial. Reason codes show the key positive and negative factors in a model's scoring decision in simple language.

Reason codes are also useful in other settings, such as healthcare, because they can provide insights into model decisions that can drive additional testing or investigation.

Automatic Model Documentation (Auto Doc)

To explain models to business users and regulators, data scientists must document the data, algorithms, and processes used to create machine learning models. Driverless AI automatic model documentation relieves the user from the time-consuming task of documenting 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. Auto Doc saves data scientists time, which can then be used to train and deploy more models.

Automatic Scoring Pipelines

H2O Driverless AI automatically generates both Python scoring and Java ultra-low latency automatic scoring pipelines. The automatic 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 ultra-fast scoring for real-time applications running on a range of devices.

Bring-Your-Own Recipes to Make Your Own AI

H2O Driverless AI can be extended by data scientists to help them make their own AI, with customizations – models, transformers, and scorers – on the platform. 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 consume custom recipes from the library to improve models. They can also develop customizations specific to their use-cases, industry, or their enterprise.

Time Series Recipes

Time-series forecasting is one of the biggest challenges for data scientists. Time-series models address key 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, incorporate data from numerous predictors, handle structured character data and high-cardinality categorical variables, and handle gaps in time series data and other missing values. Driverless AI also provides MLI for time-series models.

Enterprise-Ready

H2O Driverless AI is scalable, secure, and connects to a variety of data sources whether in the cloud or on-prem. H2O.ai are the experts in data science at scale. Driverless AI customers enjoy a full range of support, training, and expertise to assist them with their AI journey.