H₂O.ai

H2O Driverless AI

H2O Driverless AI is an automatic machine learning platform that gives you an experienced "data scientist in a box" to complement your AI Strategy.

Delivering AI at Scale

Businesses everywhere have realized that AI applications are the key to driving better customer experiences and increased profits. In every company, thousands of AI models will be required to automate and enhance workflows and accelerate innovation of new digital products. Existing machine learning systems require expensive data science talent, take months to develop and deploy a single model, and produce "black box" models that the business and regulators don't trust. H2O Driverless AI is an automatic machine learning platform that empowers data teams to scale and deliver trusted, production-ready models to meet these challenges.

Filling the Talent Gap

Data scientists are in short supply. Increasing the number of technical people who can develop production-ready AI models is key to filling the data science talent gap. With Driverless AI, novice data scientists, data engineers, domain scientists and statisticians can develop highly accurate models that are ready to deploy. Developed by expert data scientists, the Driverless AI platform performs the tasks of an advanced data scientist that can take many years to learn and perfect.

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 are driven by high-performance computing on GPU and CPU systems that allow for thousands of combinations and iterations to be tested to find the best model in minutes. Model deployment also 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 key capabilities for understanding and sharing model results including Machine Learning Interpretably (MLI) dashboards, automated model documentation and reason codes for service representatives and customers.

H₂O.ai

Key Features of H2O Driverless AI

AutoVis - Exploratory Data Analysis for Big Data

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 to see trends and possible issues such as large numbers of missing values or significant outliers that could impact modelling results.
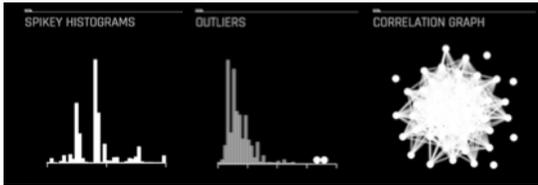

Figure 1: Sample Autoviz Charts Selected Based on Data Shape

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 employs a library of algorithms and feature transformations to automatically engineer new, high value features for a given dataset. 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)

H2O Driverless AI provides robust interpretably of machine learning techniques and results including automatically generated K-LIME, Shapley, Variable Importance, Decision Tree and Partial Dependence charts. Each chart helps to explore the modeling techniques and results more closely. These techniques are crucial for those who must explain their models to business stakeholders and regulators.
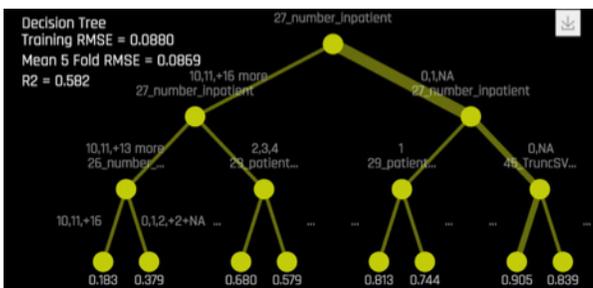

Figure 2: MLI Charts Example, Decision Tree Surrogate Model

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 and data engineers must document the data, algorithms and process 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, and the final model created. With this capability in Driverless AI, practitioners can focus more on drawing actionable insights from the models and save weeks or even months in the development, validation and deployment process.

Automatic Scoring Pipelines

H2O Driverless AI automatically generates both Python scoring pipelines and new ultra-low latency automatic scoring pipelines. The new automatic scoring pipeline is a unique technology that deploys the feature engineering and the winning machine learning model or ensemble in a highly optimized, low latency production ready Java pipeline 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.

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 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.

Enterprise Data Access and Security

Enabling AI for enterprise deployment requires data access, scalability, and security. Driverless AI support enterprise data access with connectors to on-premise and cloud data sources including AWS, Google Cloud, Azure and Snowflake. Driverless AI supports both CPU and GPU-based systems including the latest from NVIDIA and IBM Power. For user authentication, Driverless AI supports LDAP and Kerberos to provide secure access.

Enterprise Support & Training

H2O.ai is known for superior customer service and support. In the 2018 Gartner Magic Quadrant for Data Science and Machine Learning Platforms, H2O.ai received the highest score for customer service and support. H2O.ai enterprise support services include: expert 24x7 email & phone support; access to H2O data science, MLI, and algorithm experts; health checks, DevOps / production consultation; and access to online and onsite training services.

H2O.ai