

H2O AutoDoc

Automated Model Documentation



Create comprehensive, high-quality model documentation in minutes that saves time, increases productivity and improves model governance.

The Model Documentation Paradox

For many companies, model documentation is a requirement for any model to be used in the business. For other companies, model documentation is part of a data science team’s best practices. Model documentation includes how a model was created, training and test data characteristics, what alternatives were considered, how the model was evaluated, and information on model performance. Collecting and documenting this information can take a data scientist days to complete for each model. The model document needs to be comprehensive and consistent across various projects. The process of creating this documentation is tedious for the data scientist and wasteful for the business because the data scientist could be using that time to build additional models and create more value. Inconsistent or inaccurate model documentation can be an issue for model validation, governance, and regulatory compliance.

H2O AutoDoc

H2O Automated Model Documentation (AutoDoc) automatically creates model documentation for supervised learning models created in H2O-3 and Scikit-Learn. Automated documentation has been used in production in H2O Driverless AI. This industry-leading capability is now available to everyone who uses H2O-3.

Key Capabilities

- Automatic document generation in Microsoft Word (docx) or Markup (.md) formats.
- Out-of-the-box documentation template included
- Template customization available to fit with your organization’s standards and requirements
- Support for models generated in H2O-3 and Scikit-Learn
- Support for H2O-3: Deep Learning, Distributed Random Forrest, GLM, Gradient Boosted Machines, Stacked Ensembles, and XGBoost models.

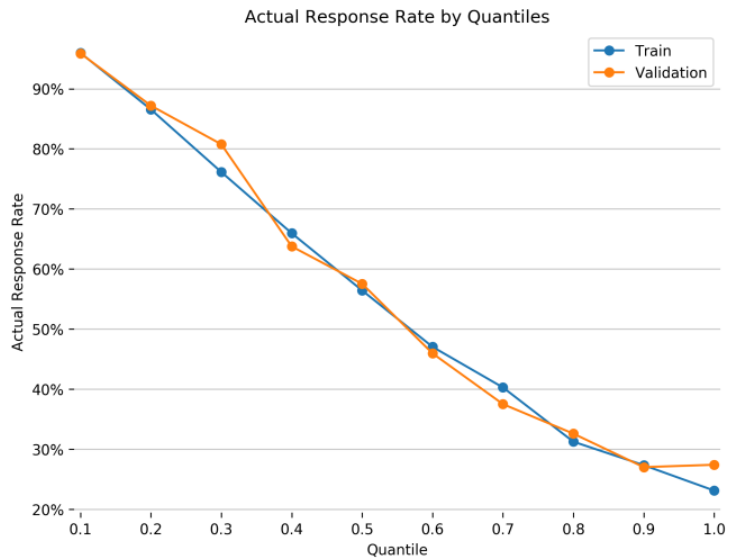


Figure 1: Example Plot - Actual Response Rate by Quantiles

Documentation Features

The Word editable document generated has these standard automated documentation template and includes:

- **Experiment Overview** to provide an overview of the modeling problem
- **System Specifications** to describe the exact configuration of the system that produced the model including the version of H2O-3 or Scikit-learn that was used.
- **Data Overview** including information on the data shape and summary statistics for each feature (numeric and categorical values)
- **Data Shift** to highlight any difference between training and validation data.
- **Validation Strategy**
- **Model Parameters and Values**
- **Common Classification or Regression Metrics**
- **Population Stability Index**
- **Prediction Statistics** for training and validation datasets
- **Feature Importance** using H2O native importance or Shapely Importance

- **Response Rate by Quantile**
- **Actual vs. Predicted Probabilities**
- **Partial Dependence Plots**
- **Alternative Models Summary** to show the other techniques and their parameters that were tested against the winning model

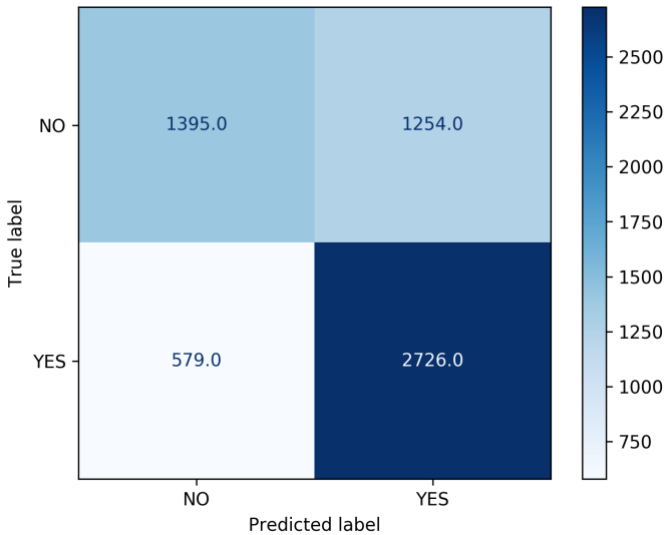


Figure 2 - Example Plot - Confusion Matrix

scorer	training	validation
AUC	0.7815	0.7793
ACCURACY	0.7127	0.716
F1	0.7498	0.7484
MCC	0.4327	0.4367
LOGLOSS	0.5567	0.5596

Figure 4: Example Table - Final Model Performance Metrics

Save Time and Money with AutoDoc

H2O AutoDoc automatically generates comprehensive model documentation in minutes using out-of-the-box or custom templates. AutoDoc saves data science team weeks of tedious work and increases data science productivity by allowing them to focus on model building. AutoDoc increases the consistency of model documentation by applying a standard template across all models and teams which is important for model governance, reproducibly, and compliance with regulations.

For more information

Visit us at www.h2o.ai

Contact us at sales@h2o.ai

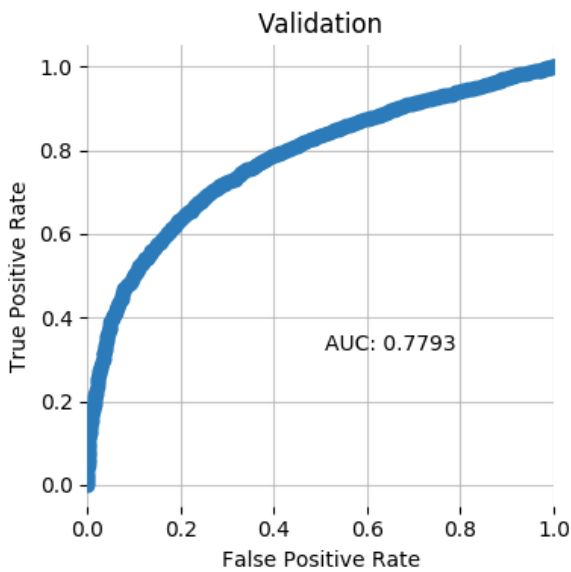


Figure 3: Example Plot - Receiver Operating Characteristic Curve

About H2O.ai

H2O.ai is the open source leader in AI and automatic machine learning with a mission to democratize AI for everyone. H2O.ai is transforming the use of AI to empower every company to be an AI company in financial services, insurance, healthcare, telco, retail, pharmaceuticals and marketing. H2O.ai is driving an open AI movement with H2O, which is used by more than 20,000 companies and hundreds of thousands of data scientists. H2O Driverless AI, an award winning and industry leading automatic machine learning platform for the enterprise, is helping data scientists across the world in every industry be more productive and deploy models in a faster, easier and cheaper way. H2O.ai partners with leading technology companies such as NVIDIA, IBM, AWS, Intel, Microsoft Azure and Google Cloud Platform and is proud of its growing customer base which includes Capital One, Nationwide Insurance, Walgreens and MarketAxess. H2O.ai believes in AI4Good with support for wildlife conservation and AI for academics. Learn more at www.H2O.ai