Open for Innovation
KNIME

KNIME H2O Driverless AI Integration User Guide

KNIME AG, Zurich, Switzerland
Version 5.3 (last updated on 2021-12-16)

Table of Contents

Introduction . . . 1
Installation . . . 2
Installing the extension . . . 2
Specifying the H2O Driverless AI license file . . . 3
Python requirement . . . 4
H2O Driverless AI Online Setup . . . 5
H2O Driverless AI usage examples . . . 6
Pushing data to H2O Driverless AI . . . 6
Reading MOJO Scoring Pipelines into KNIME . . . 9
Running H2O Driverless AI Experiments from KNIME . . . 10
Using H2O Driverless AI MOJOs for predictions . . . 13

Introduction

This guide explains how to connect KNIME Analytics Platform with H2O Driverless AI, allowing you to push tables from KNIME to H2O, run experiments from KNIME workflows, or to read in H2O Driverless AI MOJO files for use in KNIME workflows. The following sections will explain how to install the required extension and how to integrate the two solutions together.

© 2024 KNIME AG. All rights reserved.

Installation

In order to use the H2O Driverless AI nodes, you will need to install the extension in KNIME Analytics Platform, specify a H2O Driverless AI license file, and ensure KNIME is using a python environment with the correct packages.

Installing the extension

You can install the KNIME H2O Driverless AI Integration extension by drag and drop from KNIME Hub or go to *File* → *Install KNIME Extensions…* in KNIME Analytics Platform toolbar and select KNIME H2O Driverless AI Integration extension under *KNIME Partner Extensions*, as shown in Figure 1.

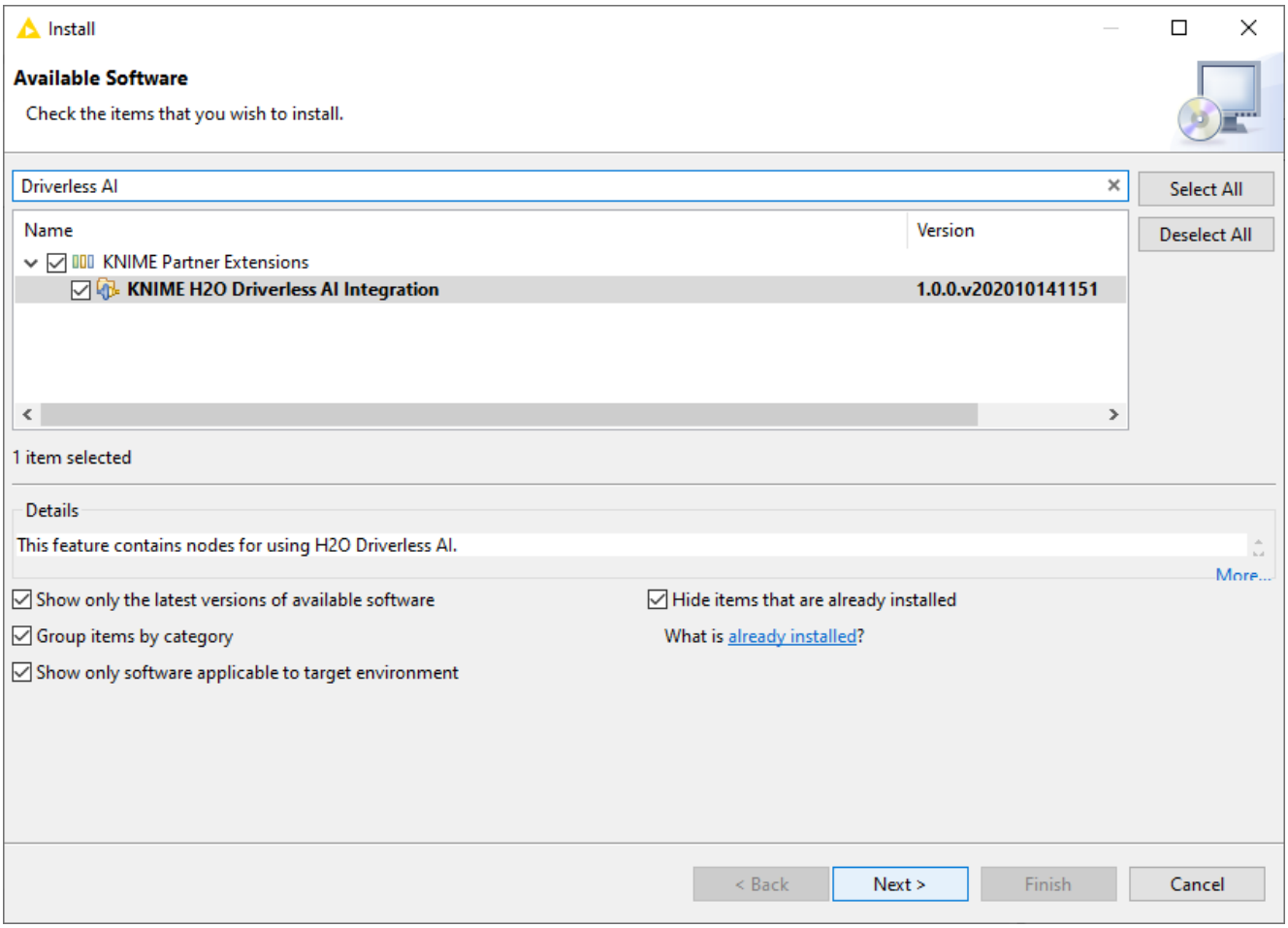

Figure 1. Installing KNIME H2O Driverless AI Integration

Once you have installed the extension, restart KNIME Analytics Platform and you should see the following nodes in the node repository under *H2O Driverless AI* as shown in Figure 2:

© 2024 KNIME AG. All rights reserved.

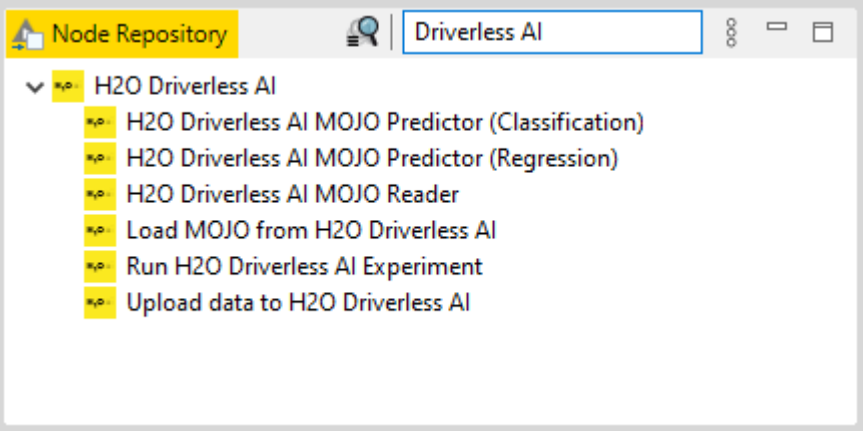

Figure 2. The H2O Driverless AI nodes in the node repository

You can now drag and drop the H2O Driverless AI nodes into your workflow and use them to integrate with the H2O Driverless AI service.

Specifying the H2O Driverless AI license file

In order to utilize the H2O Driverless AI nodes, you will need to import an H2O Driverless AI license file. There are three methods to do this:

1. Within KNIME, navigate to *File* → *Preferences* → *KNIME* → *H2O Driverless AI* and, as shown in Figure 3, upload the `.sig` file provided by H2O.

© 2024 KNIME AG. All rights reserved.

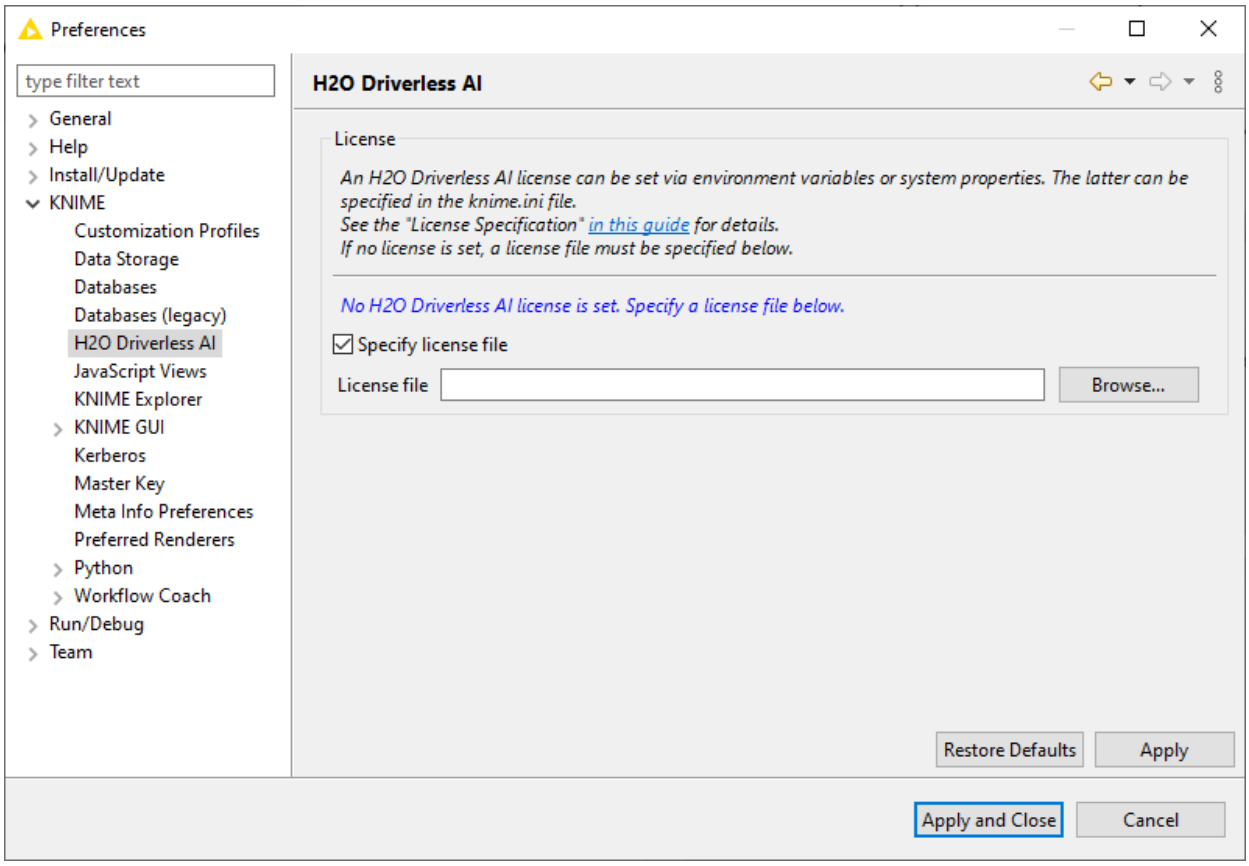

Figure 3. H2O Driverless AI preferences in KNIME Analytics Platform

2. Specify via an environment variable as described in the H2O documentation
3. Specify via a system property of the JVM as described in the H2O documentation. This system property will need to be inserted into the `knime.ini` file.

Python requirement

A working python environment is required. After setting up your `conda` environment, following instructions on the KNIME Python Integration Installation Guide, the only additional dependency required in the environment is the `driverlessai` module. This can be done via the command

```
pip install driverlessai
```

Additionally, the `.yml` script below is provided, which can be used to build a `conda` environment that will support this integration.

py36_knime_h2o-dai.yml

© 2024 KNIME AG. All rights reserved.

```
name: py36_knime_h2o-dai  # Name of the created environment
channels:                 # Repositories to search for packages
- defaults
- anaconda
- conda-forge
dependencies:         # List of packages that should be installed
- python=3.6          # Python
- pandas=0.23         # Table data structures
- jedi=0.13           # Python script autocompletion
- python-dateutil=2.7 # Date and Time utilities
- numpy=1.15          # N-dimensional arrays
- cairo=1.14          # SVG support
- pillow=5.3          # Image inputs/outputs
- matplotlib=3.0      # Plotting
- pyarrow=0.11        # Arrow serialization
- IPython=7.1         # Notebook support
- nbformat=4.4        # Notebook support
- scipy=1.1           # Notebook support
- jpype1=0.6.3        # Databases
- pip:                # Install driverless ai via pip
  - driverlessai==1.10.1 # Adapt that version if necessary
```

After downloading the `.yml` file, run the command:

```
conda env create -f py36_knime_h2o-dai.yml
```

H2O Driverless AI Online Setup

For the purpose of this guide, we will assume you are working with an instance of H2O Driverless AI. If you do not already have a license for this service, you can do so through the H2O.ai site or create a trial account. To utilize the H2O Driverless AI nodes, you will need a few pieces of information:

- The URL of your H2O Driverless AI instance
- Your username
- Your password
- Any MOJO Scoring Pipelines you would like to read into your KNIME workflows

Make sure you download the Java-compatible version of the MOJO.

© 2024 KNIME AG. All rights reserved.

H2O Driverless AI usage examples

Pushing data to H2O Driverless AI

This section explains how to configure the Upload data to H2O Driverless AI node to send KNIME tables to a H2O Driverless AI instance. An example workflow is shown in Figure 4.

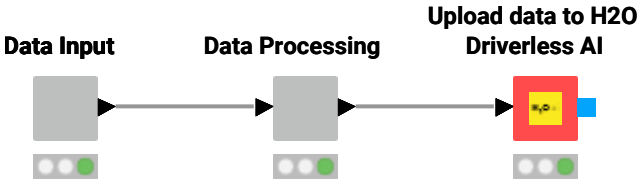

Figure 4. Example workflow to push data from KNIME Analytics Platform to H2O Driverless AI

> An example workflow about the usage of the H2O Driverless AI nodes is available on KNIME Hub.

Upload data to H2O Driverless AI node

In order to upload KNIME data tables to H2O Driverless AI, connect your workflow to the Upload data to H2O Driverless AI node. Right-click the node and select *Configure…* from the context menu. The node configuration dialog, shown in Figure 5, opens:

© 2024 KNIME AG. All rights reserved.

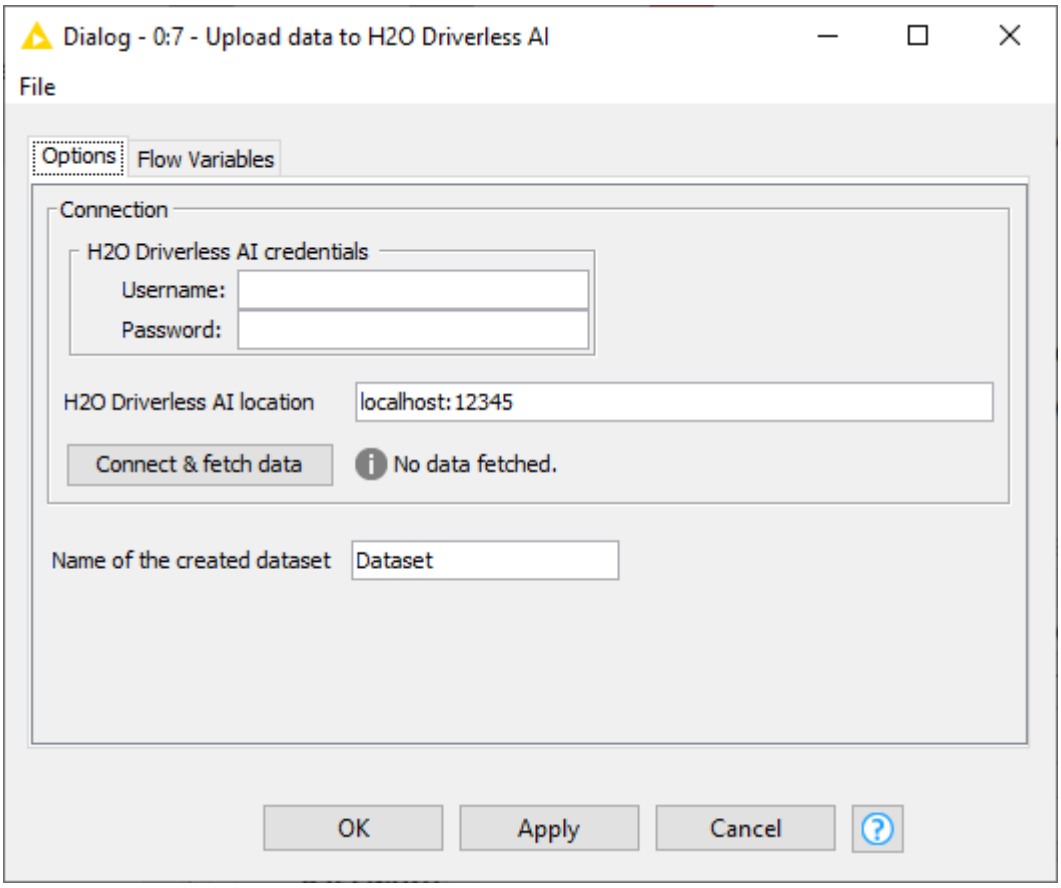

Figure 5. The Upload data to H2O Driverless AI node configuration dialog

Here, you will set the credentials, connection information, and the desired name for the dataset. Upon executing this workflow, you should see your dataset appear in the *Datasets* tab within H2O Driverless AI, as shown in Figure 6.

© 2024 KNIME AG. All rights reserved.

Figure 6. A dataset in the Datasets tab within H2O Driverless AI

Once the Upload data to H2O Driverless AI node has been executed, you can right-click and open the *Interactive View: H2O Driverless AI Experiment View*. This view, shown in Figure 7, displays your H2O Driverless AI UI from inside a KNIME view window. From here, you can interact with your Driverless AI instance and initiate an Experiment with the data you have uploaded.

© 2024 KNIME AG. All rights reserved.

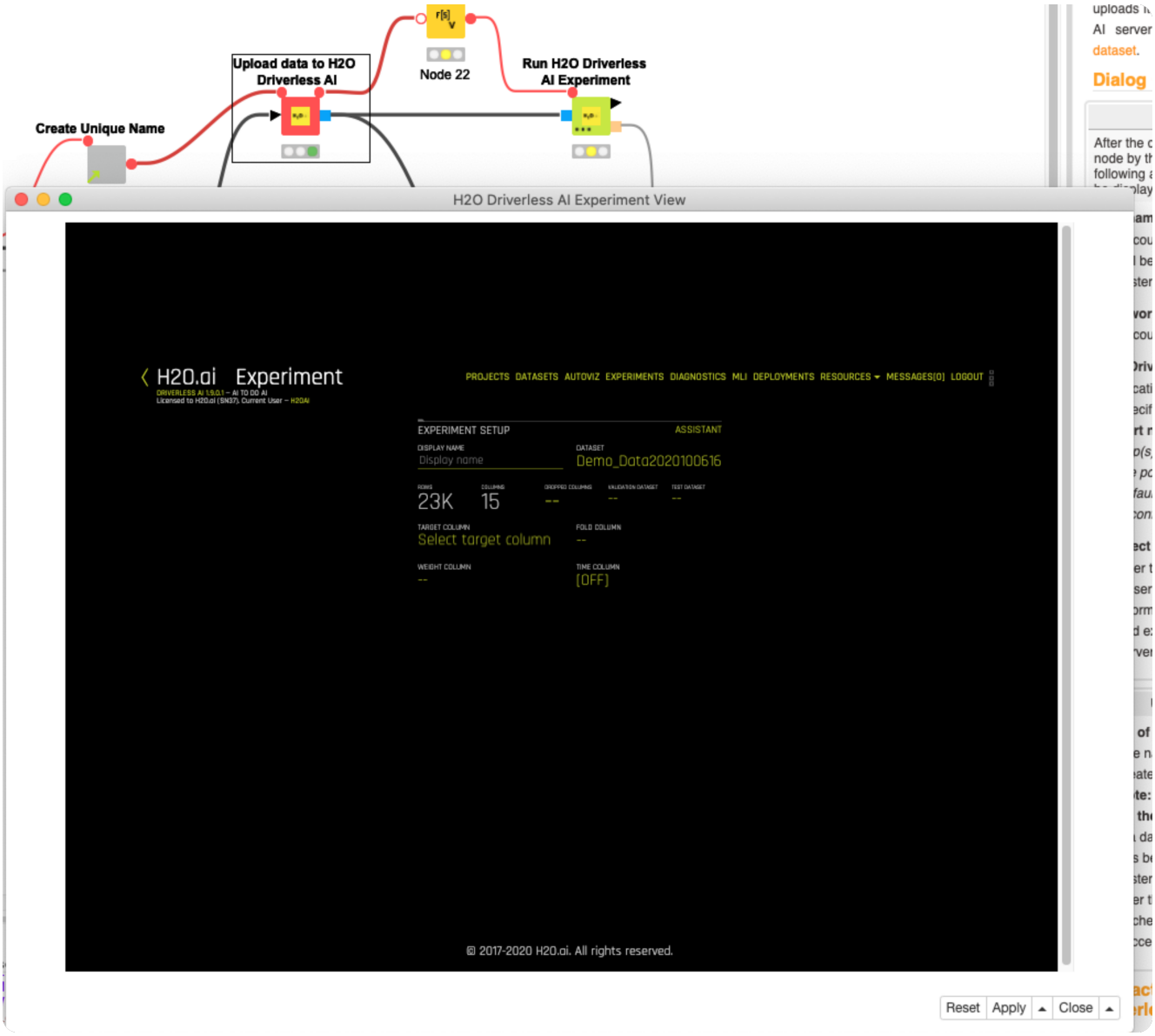

Figure 7. The H2O Driverless AI Experiment View

Reading MOJO Scoring Pipelines into KNIME

This section explains how to utilize H2O Driverless AI MOJO Scoring Pipelines in KNIME using the H2O Driverless AI MOJO Reader and Load MOJO from H2O Driverless AI nodes.

H2O Driverless AI MOJO Reader node

This node is used when you have the H2O Driverless AI MOJO file stored outside of H2O Driverless AI. When configuring the node, set where you want to read from, and point to the `pipeline.mojo` file within the `mojo-pipeline` directory.

If you would like to read the MOJO file from a remote file system, click the three little dots on the node and select *Add Filesystem Connection port*, as shown in Figure 8.

© 2024 KNIME AG. All rights reserved.

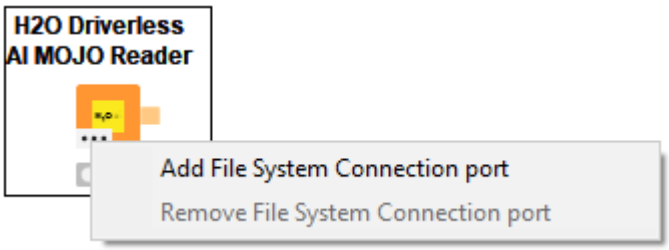

Figure 8. Add file system connection port to H2O Driverless AI MOJO Reader node

Load MOJO from H2O Driverless AI node

This node is used when you want to import a MOJO into KNIME directly from Driverless AI. Upon entering H2O Driverless AI credentials and H2O Driverless AI location URL, you have the option to click *Connect & fetch data* to populate the *Dataset* and *Experiment* fields. Choosing these values will pull the MOJO Scoring information into KNIME for use with the H2O Driverless AI MOJO Predictor nodes.

If you want a more dynamic connection in regards to the workflow, you can click the three dots on the node and click *Add Driverless AI Dataset Connection port* for use with the Upload data to H2O Driverless AI node. This node will use the connection information from the previous node to pull MOJO information from your H2O Driverless AI instance.

> In this way, the node will only return MOJO files from experiments involving the dataset uploaded in the previous node.

An example workflow is shown in Figure 9.

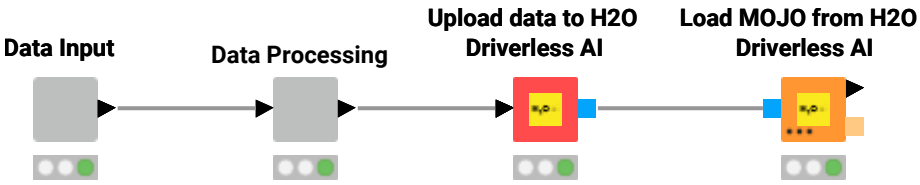

Figure 9. Example workflow to load MOJOs from H2O Driverless AI to KNIME Analytics Platform

Running H2O Driverless AI Experiments from KNIME

This section will explain how to utilize the Run H2O Driverless AI Experiment node in order to run an experiment without the need to access the H2O Driverless AI UI. An example workflow is shown in Figure 10.

© 2024 KNIME AG. All rights reserved.

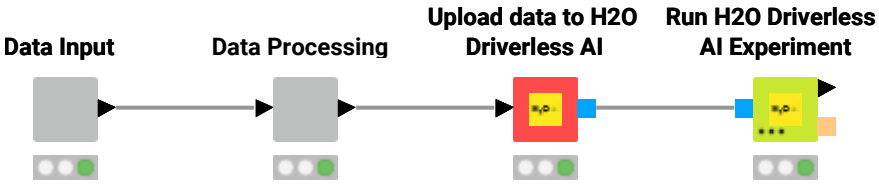

Figure 10. Example workflow to run an Experiment in H2O Driverless AI from KNIME Analytics Platform

The Run H2O Driverless AI Experiment node configuration dialog is shown in Figure 11. Enter the information about how you want your Experiment to be run and execute the node to begin the training process within H2O Driverless AI. For more information about the parameters to the Experiment, refer to the H2O Driverless AI documentation.

> It is important to note that the execution will be passed to the H2O Driverless AI instance. The node will show a running status until the results are passed back into KNIME Analytics Platform.

> The *Prediction Task* field will dictate whether the Experiment is catering to a Regression problem or a Classification problem. This is important when using the H2O Driverless AI MOJO Predictor nodes.

© 2024 KNIME AG. All rights reserved.

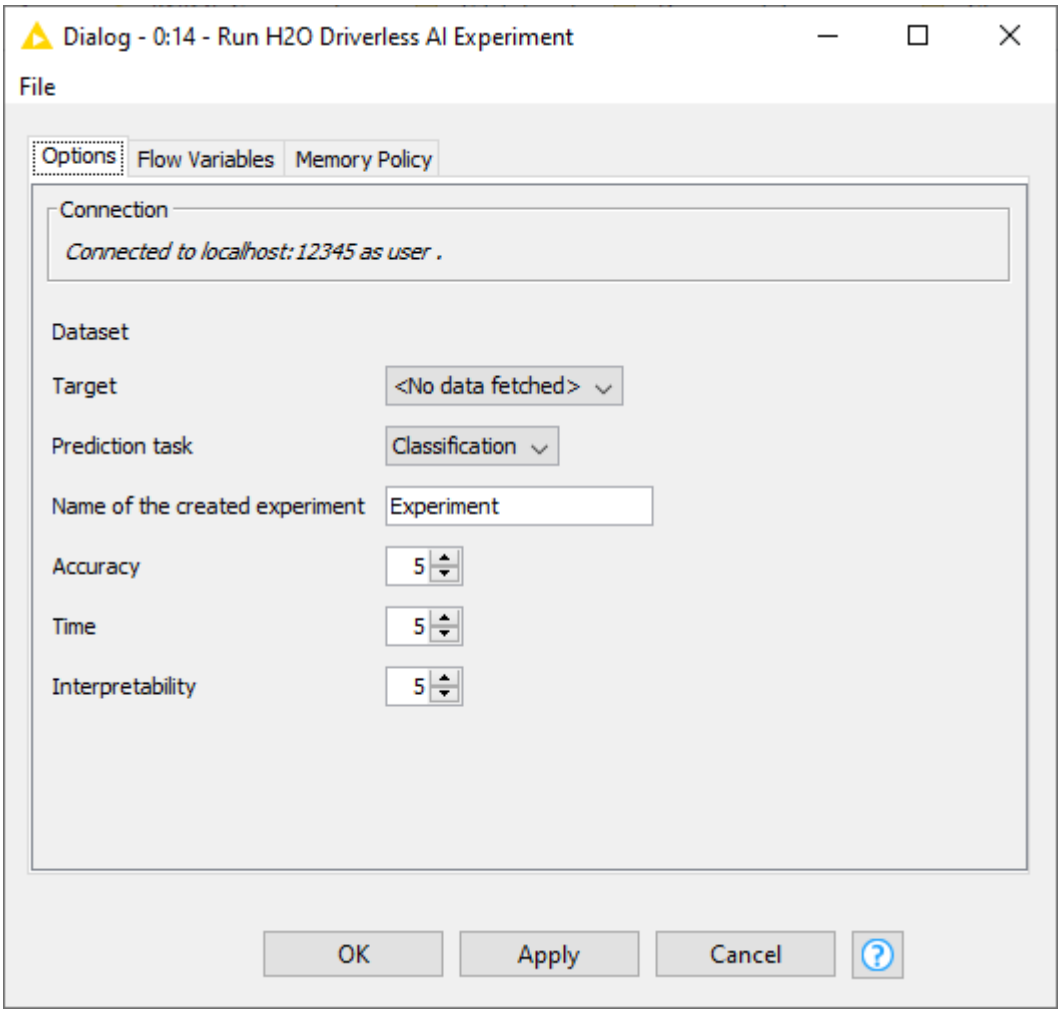

Figure 11. The Run H2O Driverless AI Experiment node configuration dialog

Upon completion, this node will output:

- The created H2O Driverless AI MOJO that contains the results of the Experiment
- An Experiment summary detailing the scoring metrics of the model created.

Once you have executed the node, you can right-click and open the *Interactive View: H2O Driverless AI Experiment View*. This view, shown in Figure 12, displays the Experiment results from the H2O Driverless AI UI. Here you can navigate the Experiment dashboard to look into the details of your Experiment.

© 2024 KNIME AG. All rights reserved.

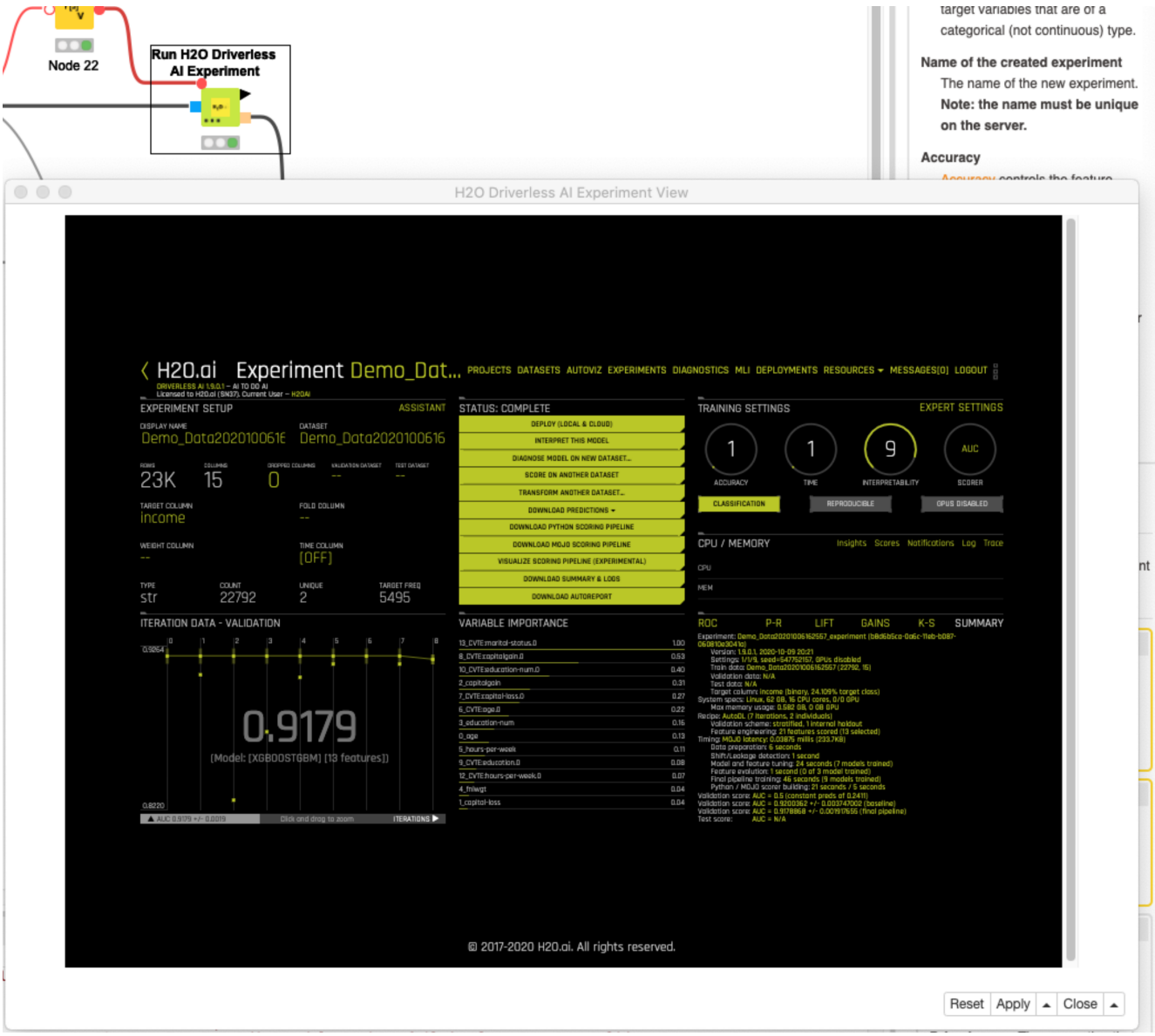

Figure 12. The H2O Driverless AI Experiment View displays the Experiment results

Using H2O Driverless AI MOJOs for predictions

This section will explain how to use the H2O Driverless AI MOJO Predictor (Regression) and H2O Driverless AI MOJO Predictor (Classification) nodes. This will allow you to use the read-into-KNIME MOJO Scoring Pipelines to make predictions on datasets.

Whether to use the Classification predictor or the Regression predictor depends on how you ran the H2O Driverless AI Experiment. The corresponding MOJO created will inherit that attribute and will only be able to be used with the matching node.

> Hovering your mouse over the MOJO port will tell you if it is passing a Classification or Regression MOJO, as shown in Figure 13.

© 2024 KNIME AG. All rights reserved.

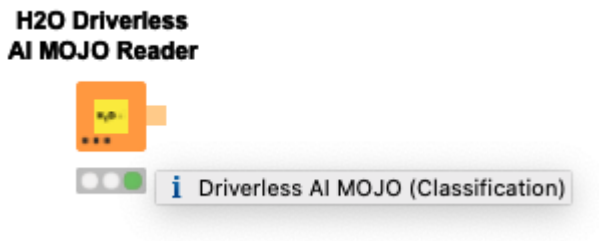

Figure 13. Information displayed hovering the mouse over MOJO port

To run predictions on your data tables, connect both a MOJO connector and a data table to the predictor node. An example workflow is shown in Figure 14.

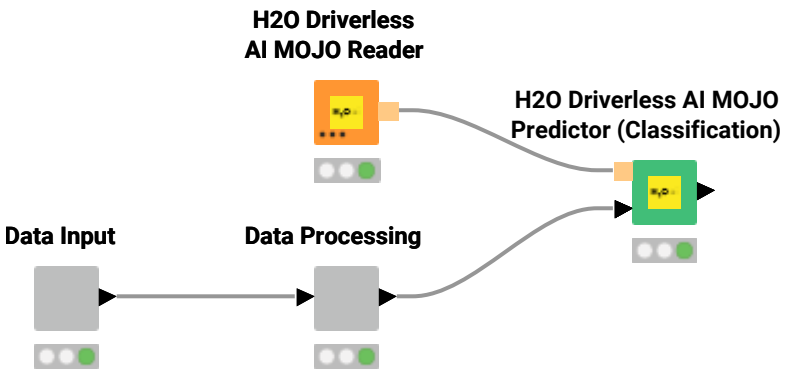

Figure 14. Example workflow to run MOJO prediction in H2O Driverless AI from KNIME Analytics Platform

> In this workflow, we use the classification version of the node because the corresponding MOJO was from a Classification Experiment.

Upon configuring the Dialog window and executing the node, it will output a table with predictions.

- For regression, the data will be appended with the predicted value for the target
- For classification, the data will be appended with columns representing probabilities of each value and, if selected, a column for the prediction (class with highest probability).

© 2024 KNIME AG. All rights reserved.

KNIME AG
Talacker 50
8001 Zurich, Switzerland
www.knime.com
info@knime.com

The KNIME® trademark and logo and OPEN FOR INNOVATION® trademark are used by KNIME AG under license from KNIME GmbH, and are registered in the United States. KNIME® is also registered in Germany.