Article

[Yuri Marx](#) · Dec 19, 2021 5m read

 Open Exchange

IntegratedML walkthrough

The InterSystems IRIS IntegratedML feature is used to get predictions and probabilities using the AutoML technique. The AutoML is a Machine Learning technology used to select the better Machine Learning algorithm/model to predict status, numbers and general results based in the past data (data used to train the AutoML model). You don't need a Data Scientist, because the AutoML it will test the most common Machine Learning algorithms and select the better algorithm to you, based in the data features analysed. See more here, in this article.

InterSystems IRIS has a built in AutoML engine, but allows to you use H2O and DataRobot too. In this article I will show to you each step to use the InterSystems AutoML engine.

Step 1 - Download the Sample app to do the exercises

1. Go to https://openexchange.intersystems.com/package/Health-Dataset

2. Clone/git pull the repo into any local directory

```
$ git clone https://github.com/yurimarx/automl-heart.git
```

3. Open a Docker terminal in this directory and run:

```
$ docker-compose build
```

4. Run the IRIS container:

```
$ docker-compose up -d
```

Step 2 - Understand the Business Scenario and the data available

The business scenario is to predict, using past data, heart diseases. The data available to do this, it is:

```
SELECT age, bp, chestPainType, cholesterol, ekgResults,
       exerciseAngina, fbsOver120, heartDisease, maxHr,
       numberOfVesselsFluro, sex, slopeOfSt, stDepression, thallium
```

```
FROM dc_data_health.HeartDisease
```

The data dictionary to the HeartDisease table is (source: https://data.world/informatics-edu/heart-disease-prediction/workspace/data-dictionary):

Column name	Type	Description
age	Integer	In years
sex	Integer	(1 = male; 0 = female)
chestPainType	Integer	Value 1: typical angina -- Value 2: atypical angina -- Value 3: non-anginal pain -- Value 4: asymptomatic
bp	Integer	Resting blood pressure (in mm Hg on admission to the hospital)
cholesterol	Integer	Serum cholestoral in mg/dl
fbsOver120	Integer	(fasting blood sugar > 120 mg/dl) (1 = true; 0 = false)
ekgResults	Integer	Resting electrocardiographic results -- Value 0: normal -- Value 1: having ST-T wave abnormality (T wave inversions and/or ST elevation or depression of > 0.05 mV) -- Value 2: showing probable or definite left ventricular hypertrophy
maxHr	Integer	Maximum heart rate achieved
exerciseAngina	Integer	Exercise induced angina (1 = yes; 0 = no)
stDepression	Double	ST depression induced by exercise relative to rest
slopeOfSt	Integer	The slope of the peak exercise ST segment -- Value 1: upsloping -- Value 2: flat -- Value 3: downsloping
numberOfVesselsFluro	Integer	Number of major vessels (0-3) colored by flourosopy
thallium	Integer	3 = normal; 6 = fixed defect; 7 = reversable defect
heartDisease	String	Value 0: < 50% diameter narrowing -- Value 1: > 50% diameter narrowing

The heartDisease it is the property that we need predict.

Step 3 - Prepare the train Data

The HeartDisease table has 270 rows. We will get 250 to train our prediction model. To do this, we will create the following view inside Management Portal > Systems Explorer > SQL:

```
CREATE VIEW automl.HeartDiseaseTrainData AS
SELECT * FROM dc_data_health.HeartDisease WHERE ID < 251
```

Step 4 - Prepare the validation Data

We will get 20 rows to validate the results of the prediction. To do this, we will create the following view inside Management Portal > Systems Explorer > SQL:

```
CREATE VIEW automl.HeartDiseaseTestData AS
SELECT * FROM dc_data_health.HeartDisease WHERE ID > 250
```

Step 5 - Create the AutoML model to predict Heart Disease

The IntegratedML allows you create an AutoML model to do predictions and probabilities (see more in
https://docs.intersystems.com/irislatest/csp/docbook/DocBook.UI.Page.cls?KEY=GIMLBASICS). To do this, we
will create the following model inside Management Portal > Systems Explorer > SQL:

```
CREATE MODEL HeartDiseaseModel PREDICTING (heartDisease) FROM automl.HeartDiseaseTrai
nData
```

The model it will get training data (learning from) from automl.HeartDiseaseTrainData view.

Step 6 - Execute the Training

Execute the training. To do this, we will execute the following SQL instruction inside Management Portal > Systems
Explorer > SQL:

```
TRAIN MODEL HeartDiseaseModel
```

Step 7 - Validate the model trained

To validate the training, we will execute the following SQL instruction inside Management Portal > Systems
Explorer > SQL:

```
VALIDATE MODEL HeartDiseaseModel FROM automl.HeartDiseaseTestData
```

We did validate the HeartDiseaseModel using testing data from the automl.HeartDiseaseTestData view.

Step 8 - Get the validation metrics

To see the validation metrics from the validation process, we will execute the following SQL instruction inside
Management Portal > Systems Explorer > SQL:

```
SELECT * FROM INFORMATION_SCHEMA_ML_VALIDATION_METRICS
```

To understand the results returned see
https://docs.intersystems.com/irislatest/csp/docbook/DocBook.UI.Page.cls?KEY=GIMLVALIDATEMODEL.

The InterSystems IRIS documentation detail the following from the validation results:

The output of VALIDATE MODEL is a set of validation metrics that is viewable in the
INFORMATION_SCHEMA_ML_VALIDATION_METRICS table.

For regression models, the following metrics are saved:

- Variance
- R-squared
- Mean squared error
- Root mean squared error

For classification models, the following metrics are saved:

- Precision — This is calculated by dividing the number of true positives by the number of predicted positives (sum of true positives and false positives).
- Recall — This is calculated by dividing the number of true positives by the number of actual positives (sum of true positives and false negatives).
- F-Measure — This is calculated by the following expression: F = 2 * (precision * recall) / (precision + recall)
- Accuracy — This is calculated by dividing the number of true positives and true negatives by the total number of rows (sum of true positives, false positives, true negatives, and false negatives) across the entire test set.

Step 9 - Execute the predictions using your new AutoML model - the last step!

To see the validation metrics from the validation process, we will execute the following SQL instruction inside Management Portal > Systems Explorer > SQL:

```
SELECT *, PREDICT(HeartDiseaseModel ) AS heartDiseasePrediction FROM automl.HeartDiseaseTestData
```

Compare the columns heartDisease (real value) and heartDiseasePrediction (the prediction value)

Enjoy!

#IntegratedML #InterSystems IRIS
Check the related application on InterSystems Open Exchange