
Artículo

[Alberto Fuentes](#) · 20 ene, 2023 Lectura de 5 min

[Open Exchange](#)

Tutorial de IntegratedML

La funcionalidad IntegratedML de InterSystems IRIS se utiliza para obtener predicciones y probabilidades mediante el uso de la técnica AutoML. AutoML es una tecnología de Machine Learning utilizada para seleccionar el mejor algoritmo/modelo de Machine Learning para predecir el estado, los números y los resultados generales basados en los datos pasados (datos utilizados para entrenar el modelo AutoML). No necesitas un científico de datos, porque AutoML probará los algoritmos de Machine Learning más comunes y seleccionará el mejor algoritmo para ti, basado en las características de los datos analizados.

InterSystems IRIS incorpora un motor AutoML, pero también permite utilizar H2O y DataRobot. En este artículo mostraré los pasos para utilizar el motor AutoML de InterSystems.

Paso 1 - Descargar la aplicación de muestra para hacer los ejercicios

1. Ve a <https://openexchange.intersystems.com/package/Health-Dataset>
2. Clonar/git pull el repositorio en cualquier directorio local

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

3. Abrir un terminal de Docker en este directorio y ejecutar:

```
$ docker-compose build
```

4. Ejecutar el contenedor de IRIS:

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

Paso 2 - Comprender el escenario y los datos

El escenario consiste en predecir, a partir de datos previos, las enfermedades cardíacas. Los datos disponibles para hacerlo, son:

```
SELECT age, bp, chestPainType, cholesterol, ekgResults,  
        exerciseAngina, fbsOver120, heartDisease, maxHr,
```

```

    numberOfVesselsFluro, sex, slopeOfSt, stDepression, thallium
FROM dc_data_health.HeartDisease

```

El diccionario de datos de la tabla HeartDisease es

(fuente: <https://data.world/informatics-edu/heart-disease-prediction/workspace/data-dictionary>):

Nombre de la columna	Tipo	Descripción
age	Número entero	En años
sex	Número entero	(1 = hombre; 0 = mujer)
chestPainType	Número entero	Valor 1: Angina típica -- Valor 2: Angina atípica -- Valor 3: Sin dolor de angina -- Valor 4: Asintomático
bp	Número entero	Presión arterial en reposo (en mm Hg al ingresar en el hospital)
cholesterol	Número entero	Colesterol sérico en mg/dl
fbsOver120	Número entero	(glucosa en sangre en ayunas > 120 mg/dl) (1 = verdadero, 0 = falso)
ekgResults	Número entero	Resultados del electrocardiograma de reposo -- Valor 0: normal -- Valor 1: con anomalía en la onda ST-T (inversiones de la onda T y/o elevación o depresión del ST de > 0,05 mV) -- Valor 2: que muestra hipertrofia ventricular izquierda probable o definida
maxHr	Número entero	Frecuencia cardíaca máxima alcanzada
exerciseAngina	Número entero	Angina inducida por el ejercicio (1 = sí, 0 = no)
stDepression	Doble	Descenso del segmento ST inducida por el ejercicio relacionado al reposo
slopeOfSt	Número entero	La pendiente del segmento ST para el ejercicio máximo -- Valor 1: pendiente ascendente -- Valor 2: plano -- Valor 3: pendiente descendente
numberOfVesselsFluro	Número entero	Número de vasos mayores (de 0 a 3) coloreados por la fluoroscopia
thallium	Número entero	3 = normal, 6 = defecto fijo, 7 = defecto reversible
heartDisease	Cadena	Valor 0: < 50% de estrechamiento del diámetro -- Valor 1: > 50% de estrechamiento del diámetro

heartDisease es la propiedad que tenemos que predecir.

Paso 3 - Preparar los datos para el entrenamiento

La tabla HeartDisease tiene 270 filas. Tomaremos 250 para entrenar nuestro modelo de predicción. Para ello, crearemos la siguiente vista dentro del Portal de administración > Explorador de sistemas > SQL:

```

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

```

Paso 4 - Preparar los datos para la validación

Tomaremos 20 filas para validar los resultados de la predicción. Para ello, crearemos la siguiente vista dentro del Portal de administración > Explorador de sistemas > SQL:

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

Paso 5 - Crear el modelo de AutoML para predecir las enfermedades cardíacas

IntegratedML permite crear un modelo de AutoML para hacer predicciones y probabilidades (más información en). Para ello, crearemos el siguiente modelo dentro de Portal de administración > Explorador de sistemas > SQL:

```
CREATE MODEL HeartDiseaseModel PREDICTING (heartDisease) FROM automl.HeartDiseaseTrainData
```

El modelo obtendrá los datos del entrenamiento (aprendizaje) de la vista automl.HeartDiseaseTrainData.

Paso 6 - Ejecutar el entrenamiento

Para hacer esto, ejecutaremos la siguiente instrucción de SQL dentro del Portal de administración > Explorador de sistemas > SQL:

```
TRAIN MODEL HeartDiseaseModel
```

Paso 7 - Validar el modelo de entrenamiento

Para validar el modelo que acabamos de entrenar, ejecutaremos la siguiente instrucción de SQL dentro del Portal de administración > Explorador de sistemas > SQL:

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

Validamos el HeartDiseaseModel utilizando datos de prueba extraídos de la vista automl.HeartDiseaseTestData.

Paso 8 - Obtener las métricas de validación

Para conocer las métricas de validación del proceso de validación, ejecutaremos la siguiente instrucción SQL dentro del Portal de administración > Explorador de sistemas > SQL:

```
SELECT * FROM INFORMATION_SCHEMA_ML_VALIDATION_METRICS
```

Para entender los resultados obtenidos, consulta:

(<https://docs.intersystems.com/irislatest/csp/docbook/DocBook.UI.Page.cls...>) La documentación de InterSystems IRIS detalla lo siguiente a partir de los resultados de validación:

La salida de VALIDATE MODEL es un conjunto de métricas de validación que se puede ver en la tabla INFORMATIONSCHEMAMLVAlIDATIONMETRICS.

Para los modelos de regresión, se guardan las siguientes métricas:

- Varianza
- R cuadrado
- Error cuadrático medio
- Error cuadrático medio de la raíz

Para los modelos de clasificación, se guardan las siguientes métricas:

- Precision — Se calcula dividiendo el número de positivos verdaderos por el número de positivos predichos (suma de positivos verdaderos y falsos positivos).
- Recall — Se calcula dividiendo el número de positivos verdaderos por el número de positivos reales (suma de positivos verdaderos y falsos negativos).
- F-Measure — Se calcula mediante la siguiente expresión: $F = 2 * (\text{precision} * \text{recall}) / (\text{precision} + \text{recall})$
- Accuracy — Se calcula dividiendo el número de verdaderos positivos y verdaderos negativos por el número total de filas (suma de verdaderos positivos, falsos positivos, verdaderos negativos y falsos negativos) en todo el conjunto de pruebas.

Paso 9 - Ejecutar las predicciones con el nuevo modelo de AutoML - ¡el último paso!

Para conocer las métricas de validación del proceso de validación, ejecutaremos la siguiente instrucción SQL dentro del Portal de administración > Explorador de sistemas > SQL:

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

Compara las columnas heartDisease (valor real) y heartDiseasePrediction (el valor de la predicción)

¡Espero que os resulte útil!

[#IntegratedML](#) [#InterSystems IRIS](#)
[Ir a la aplicación en InterSystems Open Exchange](#)

URL de fuente: <https://es.community.intersystems.com/post/tutorial-de-integratedml>