Artigo
Yuri Marx · jan 13, 2022 5min de leitura

Open Exchange

IntegratedML walkthrough - Predição de Doença cardíaca

O recurso InterSystems IRIS IntegratedML é usado para obter previsões e probabilidades usando a técnica AutoML. O AutoML é uma tecnologia de Machine Learning usada para selecionar o melhor algoritmo/modelo de Machine Learning para prever status, números e resultados gerais com base nos dados anteriores (dados usados para treinar o modelo AutoML). Você não precisa de um Cientista de Dados, pois o AutoML testará os algoritmos de Machine Learning mais comuns e selecionará o melhor algoritmo para você, com base nos recursos de dados analisados. Veja mais aqui, neste artigo.

O InterSystems IRIS possui um mecanismo AutoML integrado, mas também permite que você use H2O e DataRobot. Neste artigo, mostrarei a você cada etapa para usar o mecanismo InterSystems AutoML.

Passo 1 - Download a aplicação de exemplo e faça os exercícios

1. Vá em https://openexchange.intersystems.com/package/Health-Dataset

2. Clone/git pull do repositório para qualquer diretório local

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

3. Abra o terminal, vá no diretório do projeto e execute:

```
$ docker-compose build
```

4. Execute o IRIS no container:

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

Passo 2 - Entendendo o cenário de negócio e os dados disponíveis

O cenário de negócio é para prever, usar dados passados e assim prever doenças cardíacas. Os dados para fazer isto são:

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

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

Segue o dicionário de dados da tabela (source: https://data.world/informatics-edu/heart-disease-prediction/workspace/data-dictionary):

Column name	Type	Description
age	Integer	em anos
sex	Integer	(1 = masculino; 0 = feminino)
chestPainType	Integer	Valor 1: angina típica -- Value 2: agina atípica -- Value 3: angina não presente -- Value 4: assíntomático
bp	Integer	Pressão sanguínea (in mm Hg no momento da admissão no hospital)
cholesterol	Integer	Colesterol em mg/dl
fbsOver120	Integer	(açucar no sangue > 120 mg/dl) (1 = true; 0 = false)
ekgResults	Integer	Resultados do eletrocardiograma -- Valor 0: normal -- Valor 1: tem ST-T com ondas anormais (onda T com inversões e/ou elevação ST ou depressão de > 0.05 mV) -- Valor 2: apresenta provável hipertrofia ventricular no lado esquerdo
maxHr	Integer	Pulsação máxima do coração atingida
exerciseAngina	Integer	Angina induzida a exercícios (1 = yes; 0 = no)
stDepression	Double	Depressão ST induzida por exercício
slopeOfSt	Integer	A inclinação do segmento ST de exercício de pico -- Valor 1: ascendente -- Valor 2: regular -- Valor 3: descendente
numberOfVesselsFluro	Integer	Número de vasos principais (0-3) coloridos por fluoroscopia
thallium	Integer	3 = normal; 6 = defeito corrigido; 7 = defeito reversível
heartDisease	**String**	**Valor 0: < 50% estreitamento do diâmetro -- Valor 1: > 50% estreitamento do diâmetro**

A **heartDisease** é a propriedade que precisamos prever.

Passo 3 - Preparar o treinamento dos dados

A tabela HeartDisease tem 270 linhas. Receberemos 250 para treinar nosso modelo de previsão. Para fazer isso, vamos criar a seguinte view dentro do Management Portal > Systems Explorer > SQL:

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

Passo 4 - Preparar validação dos dados

Obteremos 20 linhas para validar os resultados da previsão. Para fazer isso, vamos criar a seguinte view dentro do Management Portal > Systems Explorer > SQL:

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

Passo 5 - Criar o modelo AutoML para a predição

O IntegratedML permite criar um modelo AutoML para fazer previsões e probabilidades (veja mais em https://docs.intersystems.com/irislatest/csp/docbook/DocBook.UI.Page.cls...). Para fazer isso, vamos criar o seguinte modelo dentro do Management Portal > Systems Explorer > SQL:

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

O modelo do qual vamos obter dados de treinamento (learning) é da view automl.HeartDiseaseTrainData.

Passo 6 - Executar o treinamento

Para fazer isso, executaremos a seguinte instrução SQL dentro do Management Portal > Systems Explorer > SQL:

```
TRAIN MODEL HeartDiseaseModel
```

Passo 7 - Validar o modelo treinado

Para validar o treinamento, executaremos a seguinte instrução SQL dentro do Management Portal > Systems Explorer > SQL:

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

Nós validamos o HeartDiseaseModel usando dados de teste da view automl.HeartDiseaseTestData.

Passo 8 - Obter as métricas de validação

Para ver as métricas de validação do processo de validação, executaremos a seguinte instrução SQL dentro do Management Portal > Systems Explorer > SQL:

```
SELECT * FROM INFORMATION_SCHEMA_ML_VALIDATION_METRICS
```

Para entender os resultados retornados, consulte:
https://docs.intersystems.com/irislatest/csp/docbook/DocBook.UI.Page.cls?KEY=GIMLVALIDATEMODEL.

A documentação do InterSystems IRIS detalha o seguinte dos resultados da validação:

A saída de VALIDATE MODEL é um conjunto de métricas de validação que podem ser visualizadas na tabela INFORMATIONSCHEMAMLVALIDATIONMETRICS.

Para modelos de regressão, as seguintes métricas são salvas:

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

Para modelo de classificação, as métricas a seguir são retornadas:

- Precision — Isso é calculado dividindo o número de verdadeiros positivos pelo número de positivos previstos (soma de verdadeiros positivos e falsos positivos).
- Recall — Isso é calculado dividindo o número de verdadeiros positivos pelo número de reais positivos (soma de verdadeiros positivos e falsos negativos).
- F-Measure — Isso é calculado pela seguinte expressão: F = 2 * (precision * recall) / (precision + recall)
- Accuracy — Isso é calculado dividindo o número de verdadeiros positivos e verdadeiros negativos pelo número total de linhas (soma de verdadeiros positivos, falsos positivos, verdadeiros negativos e falsos negativos) em todo o conjunto de teste.

Passo 9 - Execute as predições usando seu novo modelo AutoML - o último passo!

Para ver as métricas de validação do processo de validação, executaremos a seguinte instrução SQL dentro do Management Portal > Systems Explorer > SQL:

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

Compare as colunas heartDisease (valor real) e heartDiseasePrediction (o valor de predição)

Aproveite!

#IntegratedML #InterSystems IRIS
Confira o aplicativo relacionado no InterSystems Open Exchange

URL de origem:https://pt.community.intersystems.com/post/integratedml-walkthrough-predi%C3%A7%C3%A3o-de-doen%C3%A7a-card%C3%ADaca