Article
Yuri Marx · Jan 5, 2022 2m read

Open Exchange

InterSystems IRIS Open Datasets for predict important diseases

According to the WHO, The top global causes of death, in order of total number of lives lost, are associated with three broad topics (source: https://www.who.int/news-room/fact-sheets/detail/the-top-10-causes-of-death):

1. Cardiovascular (ischaemic heart disease, stroke),
2. Respiratory (chronic obstructive pulmonary disease, lower respiratory infections) and
3. Neonatal conditions – which include birth asphyxia and birth trauma, neonatal sepsis and infections, and preterm birth complications.

I created an application that's provides real data (without personal data) for some of these top 10 scenarios of diseases identified by WHO. The datasets for this application are:

- **Diabetes dataset**: data to predict diabetes diagnosis
- **Heart Disease**: data to predict heart disease
- **Kidney Disease**: data to predict kidney disease
- **Breast Cancer**: data to predict breast cancer
- **Maternal Health Risk**: data to predict maternal health risk level

To download and install the application go to https://openexchange.intersystems.com/package/Health-Dataset

Follow these instructions:

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

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

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

```
$ docker-compose build
```

3. Run the IRIS container:

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

4. Do a Select to the HeartDisease dataset:

```
SELECT
age, bp, chestPainType, cholesterol, ekgResults, exerciseAngina, fbsOver120, heartDis
ease, maxHr, numberOfVesselsFluro, sex, slopeOfSt, stDepression, thallium
FROM dc_data_health.HeartDisease
```

5. Do a Select to the Kidney Disease dataset:

```
SELECT
age, al, ane, appet, ba, bgr, bp, bu, cad, classification, dm, hemo, htn, pc, pcc, pc
v, pe, pot, rbc, rc, sc, sg, sod, su, wc
FROM dc_data_health.KidneyDisease
```

6. Do a Select to the Diabetes dataset:

```
SELECT
Outcome, age, bloodpressure, bmi, diabetespedigree, glucose, insulin, pregnancies, sk
inthickness
FROM dc_data_health.Diabetes
```

7. Do a Select to the Breast Cancer dataset:

```
SELECT
areamean, arease, areaworst, compactnessmean, compactnessse, compactnessworst, concav
epointsmean, concavepointsse, concavepointsworst, concavitymean, concavityse, concavi
tyworst, diagnosis, fractaldimensionmean, fractaldimensionse, fractaldimensionworst,
perimetermean, perimeterse, perimeterworst, radiusmean, radiusse, radiusworst, smooth
nessmean, smoothnessse, smoothnessworst, symmetrymean, symmetryse, symmetryworst, tex
turemean, texturese, textureworst
FROM dc_data_health.BreastCancer
```

8. Do a Select to the Maternal Health Risk dataset:

```
SELECT
BS, BodyTemp, DiastolicBP, HeartRate, RiskLevel, SystolicBP, age
FROM dc_data_health.MaternalHealthRisk
```

These datasets can be used into AutoML/Machine Learning applications to support breast cancer, heart disease, kidney disease and diabetes diagnostics (support only, because human doctor diagnosis is mandatory).

Enjoy!

#Data Import and Export #InterSystems IRIS
Check the related application on InterSystems Open Exchange