文章
Frank Ma · 六月 13, 2022 阅读大约需 10 分钟

Open Exchange

利用IRIS IntegratedML（一体化机器学习）来预测肾病的Web应用

肾脏疾病可以从一些医学界熟知的参数中发现。这样，为了帮助医学界和计算机系统，特别是人工智能，科学家Akshay Singh发表了一个非常有用的数据集，用于训练肾脏疾病检测/预测方面的机器学习（ML)算法。这份出版物可以在最大和最知名的ML数据库Kaggle上找到，网址是https://www.kaggle.com/datasets/akshayksingh/kidney-disease-dataset。

关于数据集

该肾脏疾病数据集有以下元数据信息（来源：https://www.kaggle.com/datasets/akshayksingh/kidney-disease-dataset):

- 它有400行，有25个特征，如红细胞、足部水肿、糖等等。
- 其目的是对病人是否患有慢性肾脏病进行分类。
- 分类是基于一个名为 "classification "（分类）的属性，属性值是 "ckd"（慢性肾脏病）或 "notckd"（不是慢性肾脏病）。
- 数据集作者对数据集进行了清洗，包括将文本映射为数字和其他一些变化。在清洗之后，数据集作者做了一些EDA（探索性数据分析），然后将数据集分为训练和测试两部分，并在上面应用模型。据观察，最初的分类结果并不令人满意。因此，数据集的作者没有放弃有Nan（非数）值的行，而是用lambda函数将其替换为每一列的模式。之后，数据集作者又将数据集分为训练集和测试集，并对其应用模型。这一次的结果更好，我们看到随机森林和决策树是表现最好的，准确率为1.0，错误分类率为0。分类的性能是通过打印混淆矩阵、分类报告和准确性来衡量的。

数据集信息 (来源: https://archive.ics.uci.edu/ml/datasets/chronickidneydisease):

我们使用以下表述来收集数据集

age - age（年龄）
bp - blood pressure（血压）
sg - specific gravity（比重）
al - albumin（白蛋白）
su - sugar（糖）
rbc - red blood cells（红血球）
pc - pus cell（脓细胞）
pcc - pus cell clumps（脓细胞团块）
ba - bacteria（细菌）
bgr - blood glucose random（血糖随机）
bu - blood urea（血尿素）
sc - serum creatinine（血清肌酐）
sod - sodium（钠）
pot - potassium（钾）
hemo - hemoglobin（血红蛋白）
pcv - packed cell volume（填充细胞体积）
wc - white blood cell count（白血球计数）
rc - red blood cell count（红细胞计数）
htn - hypertension（高血压）
dm - diabetes mellitus（糖尿病）
cad - coronary artery disease（冠状动脉疾病）
appet - appetite（食欲）

pe - pedal edema（足部水肿）
ane - anemia（贫血）
class - class（类）

属性信息 (来源: https://archive.ics.uci.edu/ml/datasets/chronickidneydisease):

我们使用24+类 = 25 （11个数字类型，14个名义类型）
1.Age 年龄（数字）
岁数
2.Blood Pressure 血压（数字）
血压单位：mm/Hg
3.Specific Gravity 比重（数字）
sg - (1.005,1.010,1.015,1.020,1.025)
4.Albumin 白蛋白（名义）
al - (0,1,2,3,4,5)
5.Sugar 糖 (数字)
su - (0,1,2,3,4,5)
6.Red Blood Cells 红血球 (名义)
rbc - (normal,abnormal) 红细胞 - (正常,异常)
7.Pus Cell 脓细胞 (名义)
pc - (normal,abnormal)（正常、异常）
8.Pus Cell clumps 脓细胞团块 (名义)
pcc - (present,notpresent) （出现、未出现）
9.Bacteria 细菌(名义)
ba - (present,notpresent) （出现、未出现）
10.Blood Glucose Random 血糖随机(数字)
bgr 单位 mgs/dl
11.Blood Urea 血尿素(数字)
bu 单位mgs/dl
12.Serum Creatinine 血清肌酸酐(数字)
sc 单位 mgs/dl
13.Sodium 钠 (数字)
sod 单位 mEq/L
14.Potassium 钾 (数字)
pot 单位 mEq/L
15.Hemoglobin 血红蛋白(数字)
hemo 单位 gms
16.Packed Cell Volume 包容细胞体积(数字)
17.White Blood Cell Count白血球计数 (数字)
wc 单位 cells/cumm
18.Red Blood Cell Count 红细胞计数(数字)
rc 单位 millions/cmm
19.Hypertension 高血压(名义)
htn - (yes,no) （是，否）
20.Diabetes Mellitus 糖尿病(名义)
dm - (yes,no) （是，否）
21.Coronary Artery Disease 冠状动脉疾病(名义)
cad - (yes,no) （是，否）
22.Appetite 食欲(名义)
appet - (good,poor) （好，差）
23.Pedal Edema 踏板水肿(名义)
pe - (yes,no) （是，否）
24.Anemia 贫血(名义)
ane - (yes,no) （是，否）
25.Class 类 （名义）
class - (ckd,notckd) （慢性肾脏病，不是慢性肾脏病）

从Kaggle获取肾脏数据

使用Health-Dataset（健康数据集）应用程序，可以把肾脏病数据从Kaggle加载到IRIS表中:
https://openexchange.intersystems.com/package/Health-Dataset.
要做到这一点，在你的module.xml项目中，设置依赖关系（Health Dataset的ModuleReference）：

Module.xml with Health Dataset application reference

预测肾脏疾病的网络前端和后端应用程序

进入Open Exchange应用程序链接 (https://openexchange.intersystems.com/package/Disease-Predictor)
并遵循以下步骤：

1. 用Clone/git 将repo拉到本地的任一目录中

```
$ git clone https://github.com/yurimarx/predict-diseases.git
```

2. 打开该目录下Docker终端，并运行:

```
$ docker-compose build
```

3. 运行IRIS容器：

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

4. 进入管理门户执行查询，训练AI模型:
   http://localhost:52773/csp/sys/exp/%25CSP.UI.Portal.SQL.Home.zen?$NAMESPACE=USER
5. 创建用于训练的VIEW：

```
CREATE VIEW KidneyDiseaseTrain AS 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. 使用view视图创建AI模型

```
CREATE MODEL KidneyDiseaseModel PREDICTING (classification) FROM KidneyDiseaseTrain
```

7. 训练模型：

```
TRAIN MODEL KidneyDiseaseModel
```

8. 访问 http://localhost:52773/disease-predictor/index.html ，使用疾病预测器的前台预测疾病，如下：

幕后工作

预测肾脏病的后端类方法

InterSystems IRIS 允许你执行SELECT，使用之前创建的模型进行预测。

Backend ClassMethod to predict Kidney Disease

现在，任何web应用都可以使用该预测并显示结果。您可以到预测疾病应用程序的前端文件夹中查看源代码。

#AI #IntegratedML #InterSystems IRIS
在 InterSystems Open Exchange 上检查相关应用程序

源URL:https://cn.community.intersystems.com/post/%E5%88%A9%E7%94%A8iris-integratedml%EF%BC%88%E4%B8%80%E4%BD%93%E5%8C%96%E6%9C%BA%E5%99%A8%E5%AD%A6%E4%B9%A0%EF%BC%89%E6%9D%A5%E9%A2%84%E6%B5%8B%E8%82%BE%E7%97%85%E7%9A%84web%E5%BA%94%E7%94%A8