Article

Yuri Marx · May 31, 2022 9m read

Open Exchange

Web App to Predict Kidney Disease using IRIS IntegratedML

Kidney Disease can be discovered from some parameters well known to the medical community. In this way, in order to help the medical community and computerized systems, especially AI, the scientist Akshay Singh published a very useful dataset for training ML algorithms in the detection/prediction of kidney disease. This publication can be found on the largest and best known data repository for ML, Kaggle at https://www.kaggle.com/datasets/akshayksingh/kidney-disease-dataset.

About the Dataset

The kidney disease dataset has the following metadata information (source: https://www.kaggle.com/datasets/akshayksingh/kidney-disease-dataset):

- It has 400 rows with 25 features like red blood cells, pedal edema, sugar,etc.
- The aim is to classify whether a patient has chronic kidney disease or not.
- The classification is based on a attribute named 'classification' which is either 'ckd'(chronic kidney disease) or 'notckd.
- The dataset author performed cleaning of the dataset which includes mapping the text to numbers and some other changes. After the cleaning, the dataset author done some EDA(Exploratory Data Analysis) and then divided the dataset int training and testing and applied the models on them. It is observed that the classification results are not much satisfying initially. So, instead of dropping the rows with Nan values the dataset author used the lambda function to replace them with mode for each column. After that the dataset author divided the dataset again into training and testing sets and applied models on them. This time the results are better and we see that the random forest and decision trees are the best performers with an accuracy of 1.0 and 0 misclassifications. The performance of the classification is measured by printing confusion matrix, classification report and accuracy.

Data Set Information (source: https://archive.ics.uci.edu/ml/datasets/chronic_kidney_disease):

We use the following representation to collect the dataset
age - age
bp - blood pressure
sg - specific gravity
al - albumin
su - sugar
rbc - red blood cells
pc - pus cell
pcc - pus cell clumps
ba - bacteria
bgr - blood glucose random
bu - blood urea
sc - serum creatinine
sod - sodium
pot - potassium
hemo - hemoglobin
pcv - packed cell volume
wc - white blood cell count
rc - red blood cell count
htn - hypertension

dm - diabetes mellitus
cad - coronary artery disease
appet - appetite
pe - pedal edema
ane - anemia
class - class

Attribute Information (source: https://archive.ics.uci.edu/ml/datasets/chronickidneydisease):

We use 24 + class = 25 (11 numeric ,14 nominal)
1.Age(numerical)
age in years
2.Blood Pressure(numerical)
bp in mm/Hg
3.Specific Gravity(nominal)
sg - (1.005,1.010,1.015,1.020,1.025)
4.Albumin(nominal)
al - (0,1,2,3,4,5)
5.Sugar(nominal)
su - (0,1,2,3,4,5)
6.Red Blood Cells(nominal)
rbc - (normal,abnormal)
7.Pus Cell (nominal)
pc - (normal,abnormal)
8.Pus Cell clumps(nominal)
pcc - (present,notpresent)
9.Bacteria(nominal)
ba - (present,notpresent)
10.Blood Glucose Random(numerical)
bgr in mgs/dl
11.Blood Urea(numerical)
bu in mgs/dl
12.Serum Creatinine(numerical)
sc in mgs/dl
13.Sodium(numerical)
sod in mEq/L
14.Potassium(numerical)
pot in mEq/L
15.Hemoglobin(numerical)
hemo in gms
16.Packed Cell Volume(numerical)
17.White Blood Cell Count(numerical)
wc in cells/cumm
18.Red Blood Cell Count(numerical)
rc in millions/cmm
19.Hypertension(nominal)
htn - (yes,no)
20.Diabetes Mellitus(nominal)
dm - (yes,no)
21.Coronary Artery Disease(nominal)
cad - (yes,no)
22.Appetite(nominal)
appet - (good,poor)
23.Pedal Edema(nominal)
pe - (yes,no)
24.Anemia(nominal)
ane - (yes,no)
25.Class (nominal)

class - (ckd,notckd)

Get the Kidney data from Kaggle

The Kidney data from Kaggle can be loaded into an IRIS table using the Health-Dataset application: https://openexchange.intersystems.com/package/Health-Dataset. To do this, from your module.xml project, set the dependency (ModuleReference for Health Dataset):

Module.xml with Health Dataset application reference

Web Frontend and Backend Application to Predict Kidney Disease

Go to Open Exchange app link (https://openexchange.intersystems.com/package/Disease-Predictor) and follow these steps:

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

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

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

```
$ docker-compose build
```

3. Run the IRIS container:

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

4. Go to Execute Query into Management Portal to train the AI model: http://localhost:52773/csp/sys/exp/%25CSP.UI.Portal.SQL.Home.zen?$NAMESPACE=USER
5. Create the VIEW used to train:

```
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. Create the AI Model using the view:

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

7. Train the model:

```
TRAIN MODEL KidneyDiseaseModel
```

8. Go to http://localhost:52773/disease-predictor/index.html to use the Disease Predictor frontend and predict

Behind the scenes

Backend ClassMethod to predict Kidney Disease

InterSystems IRIS allows you execute SELECT to predict using the previous model created.

Backend ClassMethod to predict Kidney Disease

Now, any web application can consume the prediction and show the results. See the source code into frontend folder to predict-diseases application.

#Artificial Intelligence (AI) #IntegratedML #InterSystems IRIS
Check the related application on InterSystems Open Exchange

Source URL:https://community.intersystems.com/post/web-app-predict-kidney-disease-using-iris-integratedml