HealthPrediction

neha dhake

2024-04-15

R Markdown

This package gives the prediction model for determining the probability of having diabeties based on various input parameters. These inputs are generally collected medical services by doctors and submitted to insurance companies in order to assess the health of patient and create careplan accordingly. These are generally captured by health assessments and input file for the program can be easily created from it. For this program, we need on unique id in dataset, one binary binary datapoint for patient having diabetes or not for training and building the model Once the model is build from small fraction of data, it can predict the possibility of patient having diabetes. This model is 86% accurate for dataset of 1.6L records and 1L records for training

this prorgram gives the output summary barchart for visualazation and the output.csv in data folder gives the final dataset with addition prediction column.

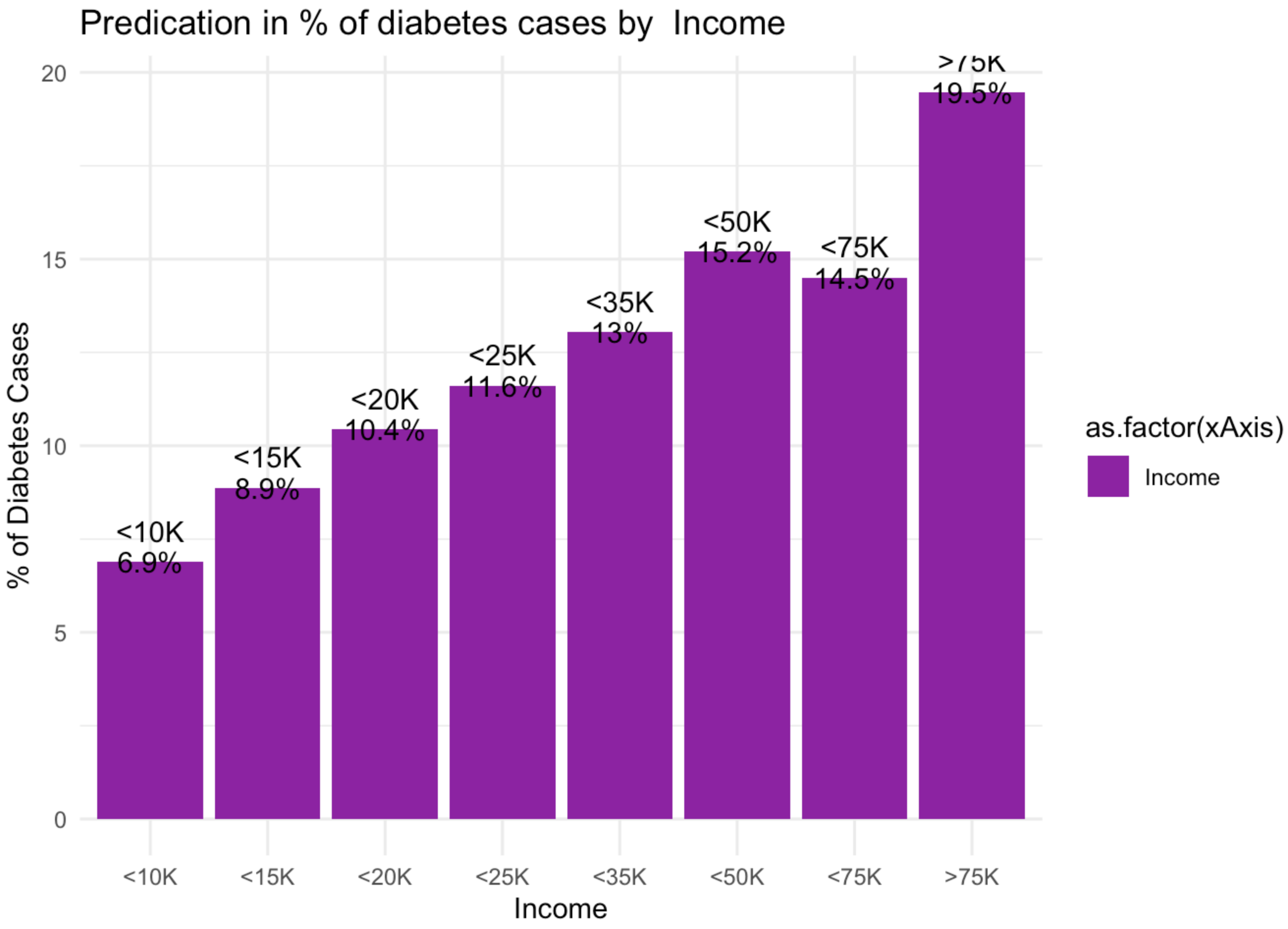
```r
script_files <- list.files("R/", full.names = TRUE, pattern = "\\.R$")

for (script in script_files) {
  print(script)
  if(script == "R//Main.R"){
    #do nothing
  } else {
    source(script)
  }
}
```

```
## [1] "R//AnalysisHelper_S4.R"
## [1] "R//AnalysisHelper.R"
## [1] "R//DatasetUtils_S4.R"
## [1] "R//DatasetUtils.R"
## [1] "R//ModelBuilder_S4.R"
## [1] "R//ModelBuilder.R"
## [1] "R//PredictionHelper_S4.R"
## [1] "R//PredictionHelper.R"
## [1] "R//PredictionManager_S4.R"
## [1] "R//PredictionManager.R"
```

```
columns <- c("ID","Diabetes_binary","HighBP","HighChol","CholCheck","BMI","Smoker","Stro
ke","HeartDiseaseorAttack","PhysActivity","Fruits","Veggies","HvyAlcoholConsump","AnyHea
lthcare","NoDocbcCost","GenHlth","MentHlth","PhysHlth","DiffWalk","Sex","Age","Educatio
n","Income")
d_column <- "Diabetes_binary"
m_columns <- c("HighBP","HighChol","CholCheck","BMI","Smoker","Stroke","HeartDiseaseorAt
tack","HvyAlcoholConsump","Age")
model_filepath <- "/Users/nehadhake/LIS6371/LIS6371-R-Programming/healthprediction/data/
Training_data.csv"
prediction_filepath <- "/Users/nehadhake/LIS6371/LIS6371-R-Programming/healthprediction/
data/Prediction_data.csv"
xAxis_column <- "Income"
xAxis_columnlabel <- c("<10K","<15K","<20K","<25K","<35K","<50K","<75K",">75K")
runprediction(model_filepath,prediction_filepath,columns,d_column,m_columns,xAxis_colum
n,xAxis_columnlabel)
```

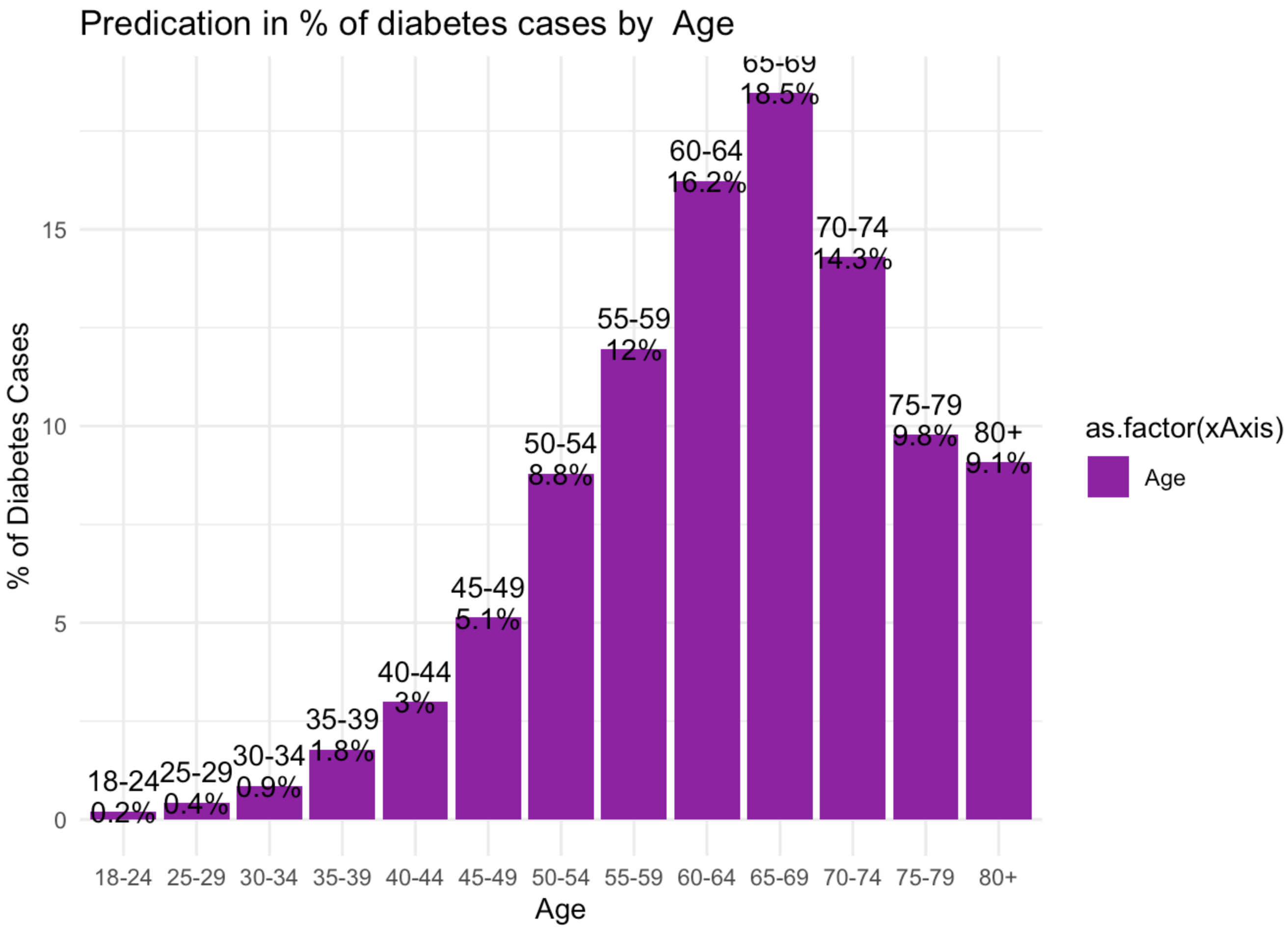
```
## [1] "Valid dataset provided"
## [1] "Valid dataset provided"
## [1] 0.8611726
## [1] "accuracy of the model is  0.861172566371681"
```



```
## [1] "please check the output files with Predicted data vs Actual Data"
```

```
xAxis_column <- "Age"
xAxis_columnlabel <- c("18-24","25-29","30-34","35-39","40-44","45-49","50-54","55-5
9","60-64","65-69","70-74","75-79","80+")
runprediction(model_filepath,prediction_filepath,columns,d_column,m_columns,xAxis_colum
n,xAxis_columnlabel)
```

```
## [1] "Valid dataset provided"
## [1] "Valid dataset provided"
## [1] 0.8611726
## [1] "accuracy of the model is  0.861172566371681"
```



```
## [1] "please check the output files with Predicted data vs Actual Data"
```