NAME : *Hariprasad K K*

REG NO : *19BCE7079*

COURSE : *DATA_ANALYTICS*

TOPIC : *Outlier Detection*

DATE : *07-11-2021*

library(devtools)

```
>
> library(devtools)
Loading required package: usethis
```

install_github("mayer79/outForest", subdir = "release/outForest")

```
> install_github("mayer79/outForest", subdir = "release/outForest")
WARNING: Rtools is required to build R packages, but is not currently ins
talled.

Please download and install Rtools 4.0 from https://cran.r-project.org/bi
n/windows/Rtools/.
Downloading GitHub repo mayer79/outForest@HEAD
Installing 4 packages: RcppEigen, FNN, ranger, missRanger
Installing packages into 'C:/Users/HP/OneDrive/Documents/R/win-library/4.
1'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/RcppEigen_0.
3.3.9.1.zip'
Content type 'application/zip' length 2870028 bytes (2.7 MB)
downloaded 2.7 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/FNN_1.1.3.zi
p'
Content type 'application/zip' length 804434 bytes (785 KB)
downloaded 785 KB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/ranger_0.13.
1.zip'
Content type 'application/zip' length 1287097 bytes (1.2 MB)
downloaded 1.2 MB

trying URL 'https://cran.rstudio.com/bin/windows/contrib/4.1/missRanger_
2.1.3.zip'
Content type 'application/zip' length 77517 bytes (75 KB)
downloaded 75 KB

package 'RcppEigen' successfully unpacked and MD5 sums checked
package 'FNN' successfully unpacked and MD5 sums checked
package 'ranger' successfully unpacked and MD5 sums checked
package 'missRanger' successfully unpacked and MD5 sums checked

The downloaded binary packages are in
        C:\Users\HP\AppData\Local\Temp\RtmpmO9jTw\downloaded_packages
WARNING: Rtools is required to build R packages, but is not currently ins
talled.

Please download and install Rtools 4.0 from https://cran.r-project.org/bi
n/windows/Rtools/.
√  checking for file 'C:\Users\HP\AppData\Local\Temp\RtmpmO9jTw\remotes44
b81a8d7447\mayer79-outForest-df8f342\release\outForest/DESCRIPTION' (717m
s)
-  preparing 'outForest': (931ms)
√  checking DESCRIPTION meta-information ...
-  checking for LF line-endings in source and make files and shell script
s
   checking for empty or unneeded directories
```

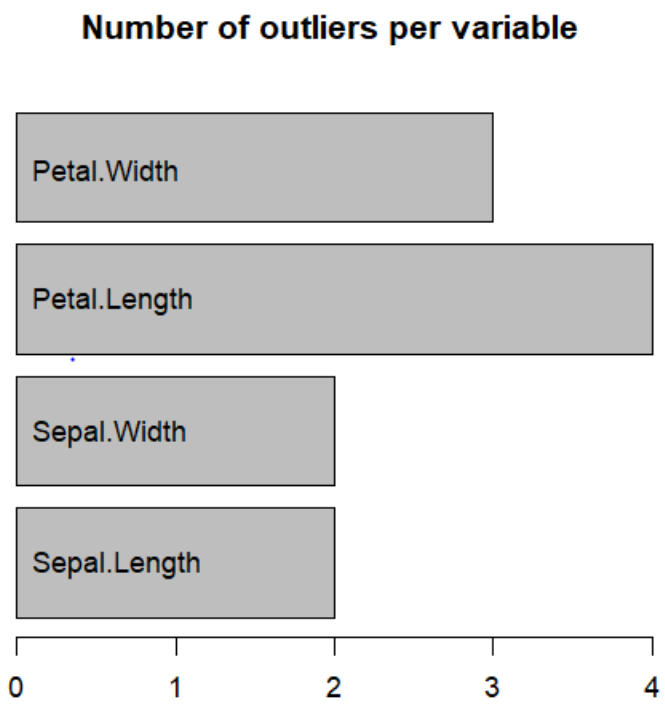
```
library(outForest)
set.seed(3)
head(irisWithOutliers <- generateOutliers(iris, p = 0.02))
```

```
> library(outForest)
> set.seed(3)
> head(irisWithOutliers <- generateOutliers(iris, p = 0.02))
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1    3.500000          1.4         0.2  setosa
2          4.9    3.000000          1.4         0.2  setosa
3          4.7    3.200000          1.3         0.2  setosa
4          4.6    3.100000          1.5         0.2  setosa
5          5.0   -3.744405          1.4         0.2  setosa
6          5.4    3.900000          1.7         0.4  setosa
>
```

(out <- outForest(irisWithOutliers, allow_predictions = TRUE))

```
> (out <- outForest(irisWithOutliers, allow_predictions = TRUE))

Outlier identification by random forests

  Variables to check:           Sepal.Length, Sepal.Width, Petal.Length,
 Petal.Width
  Variables used to check:      Sepal.Length, Sepal.Width, Petal.Length,
 Petal.Width, Species

  Checking: Sepal.Length  Sepal.Width  Petal.Length  Petal.Width  I am a
 object of class(es) outForest and list

The following number of outliers have been identified:

             Number of outliers
Sepal.Length                  2
Sepal.Width                   2
Petal.Length                  4
Petal.Width                   3
>
```

plot(out)



head(outliers(out))

```
> head(outliers(out))
   row          col  observed predicted      rmse     score threshold replacement
3    5  Sepal.Width -3.744405  3.298493 0.7810172 -9.017596         3         2.8
1   20 Sepal.Length 10.164017  5.141093 0.6750468  7.440852         3         5.4
11 138  Petal.Width  4.721186  2.113464 0.3712539  7.024092         3         2.1
9   68  Petal.Width -1.188913  1.305339 0.3712539 -6.718452         3         1.2
4  137  Sepal.Width  8.054524  2.861445 0.7810172  6.649122         3         2.9
5   15 Petal.Length  6.885277  1.875646 0.7767877  6.449163         3         1.3
> |
```

head(Data(out))

```
> head(Data(out))
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
2          4.9         3.0          1.4         0.2  setosa
3          4.7         3.2          1.3         0.2  setosa
4          4.6         3.1          1.5         0.2  setosa
5          5.0         2.8          1.4         0.2  setosa
6          5.4         3.9          1.7         0.4  setosa
> |
```

```
iris1 <- iris[1, ]az
iris1
```

```
> iris1 <- iris[1, ]
> iris1
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          5.1         3.5          1.4         0.2  setosa
> |
```

```
iris1$Sepal.Length <- -1
pred <- predict(out, newdata = iris1)
pred
```

```
> iris1$Sepal.Length <- -1
> pred <- predict(out, newdata = iris1)
> pred
I am an object of class(es) outForest and list

The following number of outliers have been identified:

             Number of outliers
Sepal.Length                  1
Sepal.Width                   0
Petal.Length                  0
Petal.Width                   0
> |
```

```
outliers(pred)
```

```
> outliers(pred)
  row          col observed predicted      rmse    score threshold replacement
1   1 Sepal.Length       -1  4.960069 0.6750468 -8.82912         3         6.4
> |
```

```
Data(pred)
```

```
> Data(pred)
  Sepal.Length Sepal.Width Petal.Length Petal.Width Species
1          6.4         3.5          1.4         0.2  setosa
> |
```