R Notebook

Code

Hide

```
install.packages("caret")
```

```
Installing package into '/Users/pulkitbatra/Library/R/arm64/4.3/library'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/src/contrib/caret_6.0-94.tar.gz'
Content type 'application/x-gzip' length 2274203 bytes (2.2 MB)
==================================================
downloaded 2.2 MB

* installing *source* package 'caret' ...
** package 'caret' successfully unpacked and MD5 sums checked
** using staged installation
** libs
using C compiler: 'Apple clang version 15.0.0 (clang-1500.0.40.1)'
using SDK: 'MacOSX14.2.sdk'
```

```
clang -I"/opt/homebrew/Cellar/r/4.3.2/lib/R/include" -DNDEBUG   -I/opt/homebrew/opt/gettext/include -I/opt/homebrew/opt/readline/include -I/opt/homebrew/opt/xz/include -I/opt/homebrew/include    -fPIC  -g -O2  -c caret.c -o caret.o
clang -dynamiclib -Wl,-headerpad_max_install_names -undefined dynamic_lookup -L/opt/homebrew/Cellar/r/4.3.2/lib/R/lib -L/opt/homebrew/opt/gettext/lib -L/opt/homebrew/opt/readline/lib -L/opt/homebrew/opt/xz/lib -L/opt/homebrew/lib -o caret.so caret.o -L/opt/homebrew/Cellar/r/4.3.2/lib/R/lib -lR -lintl -Wl,-framework -Wl,CoreFoundation
```

```
installing to /Users/pulkitbatra/Library/R/arm64/4.3/library/00LOCK-caret/00new/caret/libs
** R
** data
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
** building package indices
** installing vignettes
** testing if installed package can be loaded from temporary location
** checking absolute paths in shared objects and dynamic libraries
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (caret)

The downloaded source packages are in
    '/private/var/folders/gs/jr7fg_pj3kdbfx9sj3vfs7680000gn/T/RtmpCXxqzT/downloaded_packages'
```

Hide

```
install.packages("ggplot2")
```

```
Installing package into '/Users/pulkitbatra/Library/R/arm64/4.3/library'
(as 'lib' is unspecified)
trying URL 'https://cran.rstudio.com/src/contrib/ggplot2_3.4.4.tar.gz'
Content type 'application/x-gzip' length 3159578 bytes (3.0 MB)
==================================================
downloaded 3.0 MB

* installing *source* package 'ggplot2' ...
** package 'ggplot2' successfully unpacked and MD5 sums checked
** using staged installation
** R
** data
*** moving datasets to lazyload DB
** inst
** byte-compile and prepare package for lazy loading
** help
*** installing help indices
*** copying figures
** building package indices
** installing vignettes
** testing if installed package can be loaded from temporary location
** testing if installed package can be loaded from final location
** testing if installed package keeps a record of temporary installation path
* DONE (ggplot2)

The downloaded source packages are in
    '/private/var/folders/gs/jr7fg_pj3kdbfx9sj3vfs7680000gn/T/RtmpCXxqzT/downloaded_packages'
```

Hide

```
library(caret)
library(randomForest)
library(ggplot2)
```

Hide

```
# Load mtcars dataset
data(mtcars)

# Explore the dataset
str(mtcars)
```

```
'data.frame':   32 obs. of  11 variables:
 $ mpg : num  21 21 22.8 21.4 18.7 18.1 14.3 24.4 22.8 19.2 ...
 $ cyl : num  6 6 4 6 8 6 8 4 4 6 ...
 $ disp: num  160 160 108 258 360 ...
 $ hp  : num  110 110 93 110 175 105 245 62 95 123 ...
 $ drat: num  3.9 3.9 3.85 3.08 3.15 2.76 3.21 3.69 3.92 3.92 ...
 $ wt  : num  2.62 2.88 2.32 3.21 3.44 ...
 $ qsec: num  16.5 17 18.6 19.4 17 ...
 $ vs  : num  0 0 1 1 0 1 0 1 1 1 ...
 $ am  : num  1 1 1 0 0 0 0 0 0 0 ...
 $ gear: num  4 4 4 3 3 3 3 4 4 4 ...
 $ carb: num  4 4 1 1 2 1 4 2 2 4 ...
```

Hide

```
set.seed(123)
indices <- createDataPartition(mtcars$mpg, p = 0.7, list = FALSE)
train_data <- mtcars[indices, ]
test_data <- mtcars[-indices, ]
```

Hide

```
rf_model <- randomForest(mpg ~ ., data = train_data, ntree = 100)
```

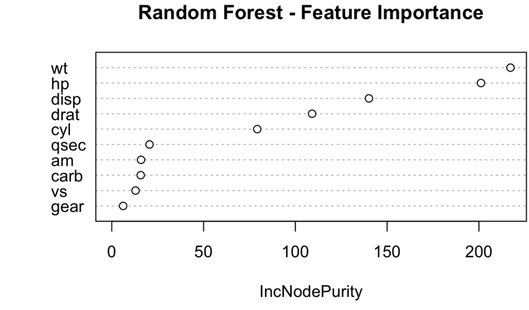
Hide

```
importance <- importance(rf_model)
print(importance)
```

```
     IncNodePurity
cyl       79.24559
disp     140.08452
hp       201.12318
drat     109.11472
wt       217.19134
qsec      20.53858
vs        12.94892
am        15.93240
gear       6.17807
carb      15.76233
```

Hide

```
# Plot feature importance
varImpPlot(rf_model, main = "Random Forest - Feature Importance")
```



Hide

```
# Predictions on the test set
predictions <- predict(rf_model, test_data)
```

Hide

```
# Scatter plot of predicted vs actual mpg
ggplot() +
  geom_point(aes(x = test_data$mpg, y = predictions), color = "blue") +
  geom_abline(intercept = 0, slope = 1, color = "red", linetype = "dashed") +
  ggtitle("Scatter Plot of Actual vs Predicted mpg") +
  xlab("Actual mpg") +
  ylab("Predicted mpg") +
  theme_minimal()
```