



# University of St.Gallen

## Forests

**University of St. Gallen**

School of Management, Economics, Law,  
Social Sciences, International Affairs  
and Computer Science

### **Assignment 3**

Data Analytics I: Predictive Econometrics  
Prof. Jana Mareckova

submitted by

**Cyril Janak, 16-611-287**  
**Jonas Husmann, 16-610-917**  
**Niklas Kampe, 16-611-618**  
**Robin Scherrer, 18-617-969**

15.12.2021

# Contents

Requirements	1
Exercise 3	1

## Requirements

To solve the following tasks, the required libraries and the data sets are loaded first.

```
library(grf)
library(DiagrammeR)
library(glmnet)

browser_2006 <- read.csv(file = "GHA/browser_2006.csv")
browser_new <- read.csv(file = "GHA/browser_new.csv")
```

## Exercise 3

To find the best two linear predictors, lasso was used. First, a 5-fold cross validation was performed to determine the optimal lambda parameter. Then, the largest coefficients of the lasso model in terms of their absolute value were determined, since they have the greatest influence on the prediction. The best two linear predictors are therefore the two websites *staples.com* and *officedepot.com*.

```
lasso.cv <- cv.glmnet(x = as.matrix(browser_2006[!names(browser_2006) %in%
                                                                    c("id", "spend")]),
                    y = browser_2006$spend,
                    type.measure = "mse",
                    family = "gaussian",
                    nfolds = 5,
                    alpha = 1)

coef.lasso.cv <- coef(lasso.cv, s = "lambda.min")
(best.lin.pred <- data.frame(name = coef.lasso.cv@Dimnames[[1]][coef.lasso.cv@i + 1],
                            coefficient = coef.lasso.cv@x) %>%
  .[order(abs(.$coefficient), decreasing = T),] %>%
  .[1:2,])
```

```
##           name coefficient
## 11    staples.com    3516.466
## 10 officedepot.com    2343.538
```