

# Woche 8

## Aufgabe 1 Titanic Challenge

Ziel der Challenge ist es vorherzusagen, ob ein Passagier das Titanic-Unglück überlebt hat, dies ist in der Variable `survived` codiert. Sie erhalten einen Teil der Daten zum Trainieren (`titanic3_train.csv`) und müssen dann für das Testset (`titanic3_test.csv`) vorhersagen, ob diese Person überlebt hat.

- a) Laden Sie die Trainings Daten `titanic3_train.csv` und die Testdaten `titanic3_test.csv`

```
train <- read.csv(file.path(baseDir, "titanic3_train.csv"), sep=";")
test  <- read.csv(file.path(baseDir, "titanic3_test.csv"), sep=";")
```

- b) Suchen Sie auf den Trainingsdaten nach dem besten Klassifikator. Verwenden Sie so viele Features oder kombinationen wie Sie wollen. Zum Beispiel:

```
klass = glm(survived ~ sex + age, data=train, family = 'binomial')
```

Mit Lösung können sie allerdings nicht gewinnen.

- c) Sagen Sie auf dem Testset vorher, ob ein Passagier überlebt hat und erzeugen Sie eine Datei mit Ihren Vorsagen der Form:

```
key;value
1;0
2;1
```

dabei steht 1 für survived, 0 für nicht. Dannach können Sie diese Datei auf <http://srv-lab-t-864/submission/Titanic/> hochladen und sehen wir gut Sie diese Daten vorher gesagt haben. Der Teilnehmer oder dass Team mit der besten Vorhersagen bekommt in der nächsten Praktikumsstunde eine Tafel Schokolade, sie können beliebig viele Vorhersagen machen.

```
pred = predict(klass, test)
df = data.frame(key=test$id, value=ifelse(pred > 0.5, 1, 0))
df[is.na(df)]=0
write.table(x=df, file = '/tmp/predictions.csv', sep=';', row.names = FALSE)
```