

SmartFly: Predict probability of delay of scheduled flights

Cindy Lamm

18:21, Friday 16th January, 2015

Load prepared data from the previous step "Prepare Data For Prediction"

```
rm(list=ls()) #clear memory
load("../04_prepare_data_for_prediction/rfPredictData.RData")
```

and load the estimated model:

```
load("../03_train_model/delayRf.RData")
```

```
library(randomForest)
# formula is_delayed ~ . - id will take care for not using id for prediction
predictionDelayRf <- predict(delayRf, rfPredictData, type="prob")
```

Save the prediction result in R format:

```
save(predictionDelayRf, file="../05_predict_delay_proba/rfPrediction.Rdata")
```

```
## Warning: Number of predicted results (532978) does not coincide with the number of scheduled
flights (566376).
```

Write the prediction result as requested into csv:

```
o <- order(predictionDelayRf[,2], decreasing=TRUE)
idIdx <- which(names(rfPredictData) == "id")
idsOrderedByDelayProba <- rfPredictData[o,idIdx]
write.table(idsOrderedByDelayProba, file="../out/problem1.csv",
            quote=FALSE, col.names=FALSE, row.names=FALSE)
```

The end!