

# Evaluate testing data (regression) - xgboost

*John Doe*

*2019-01-30*

## Contents

0. Load Data . . . . .	1
1. Scores . . . . .	2
correlation . . . . .	2
2. Important Features . . . . .	4
3. Hyper-parameters . . . . .	9

Labels: os\_time

```
## user input
project_home <- "~/EVE/examples"
project_name <- "xgboost_regression_outCV_test"
```

## 0. Load Data

## 300 of samples were used

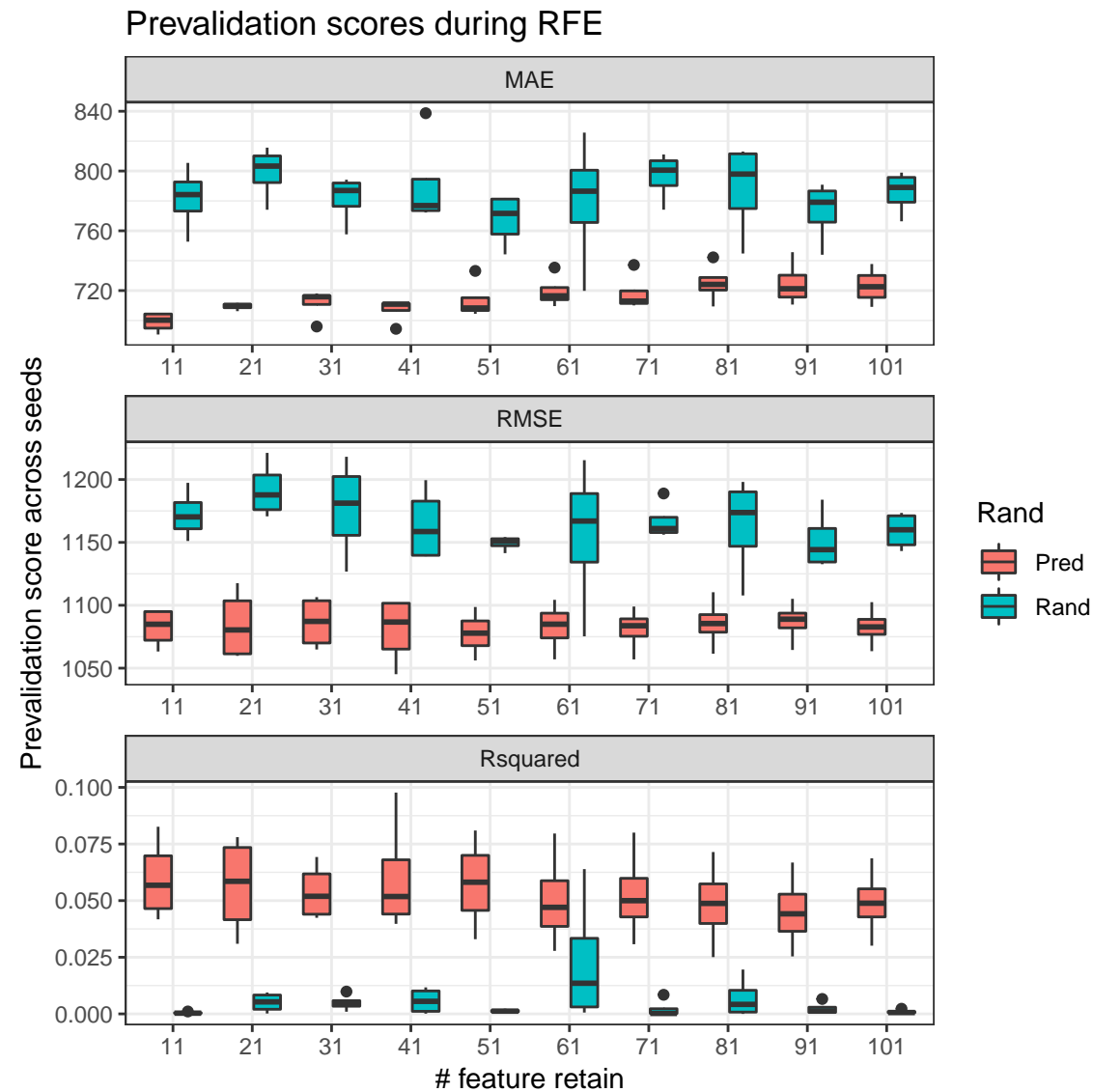
## 101 of full features

## 4 runs, each run contains 3 CVs.

## os\_time :

```
##      Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
##      0.0   182.8   480.0   889.4  1221.2   7125.0
```

## 1. Scores

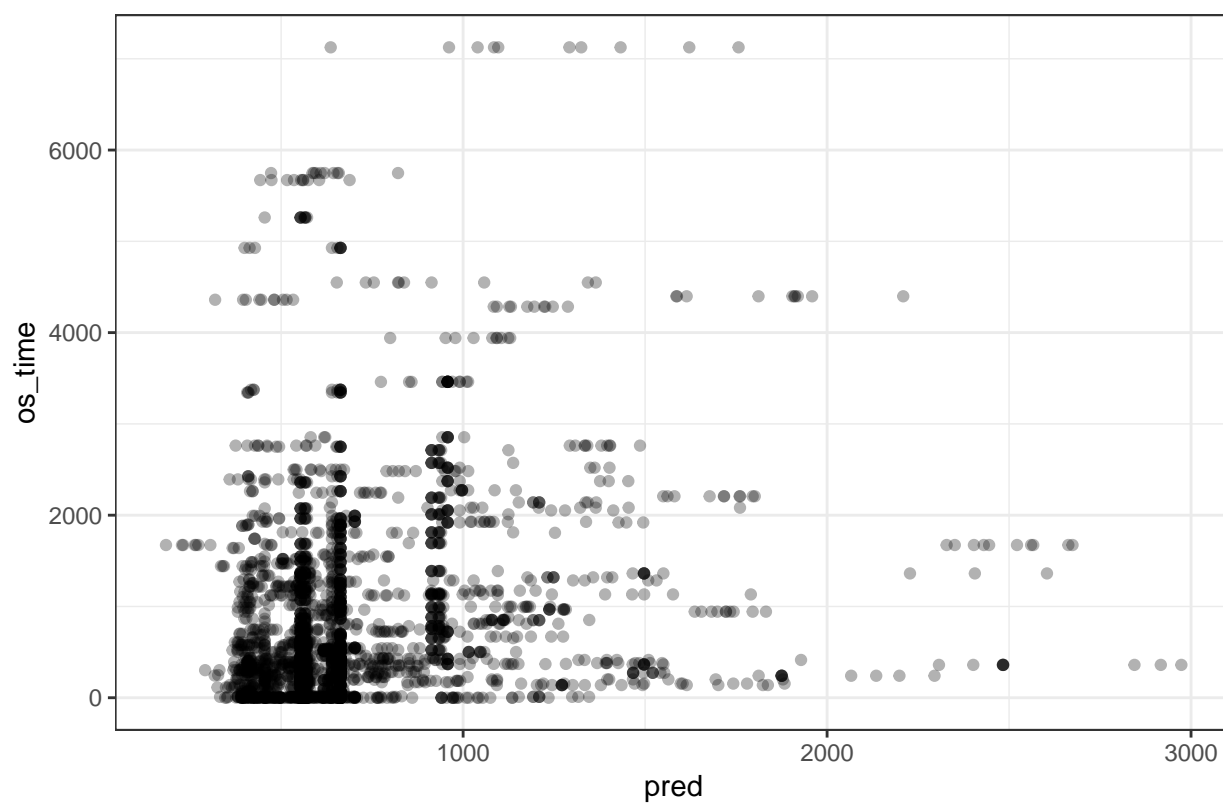


metrics	size.max	median.max	size.min	median.min
MAE	81	724.110	11	700.205
RMSE	91	1088.911	51	1077.821
Rsquared	21	0.059	91	0.044

## correlation

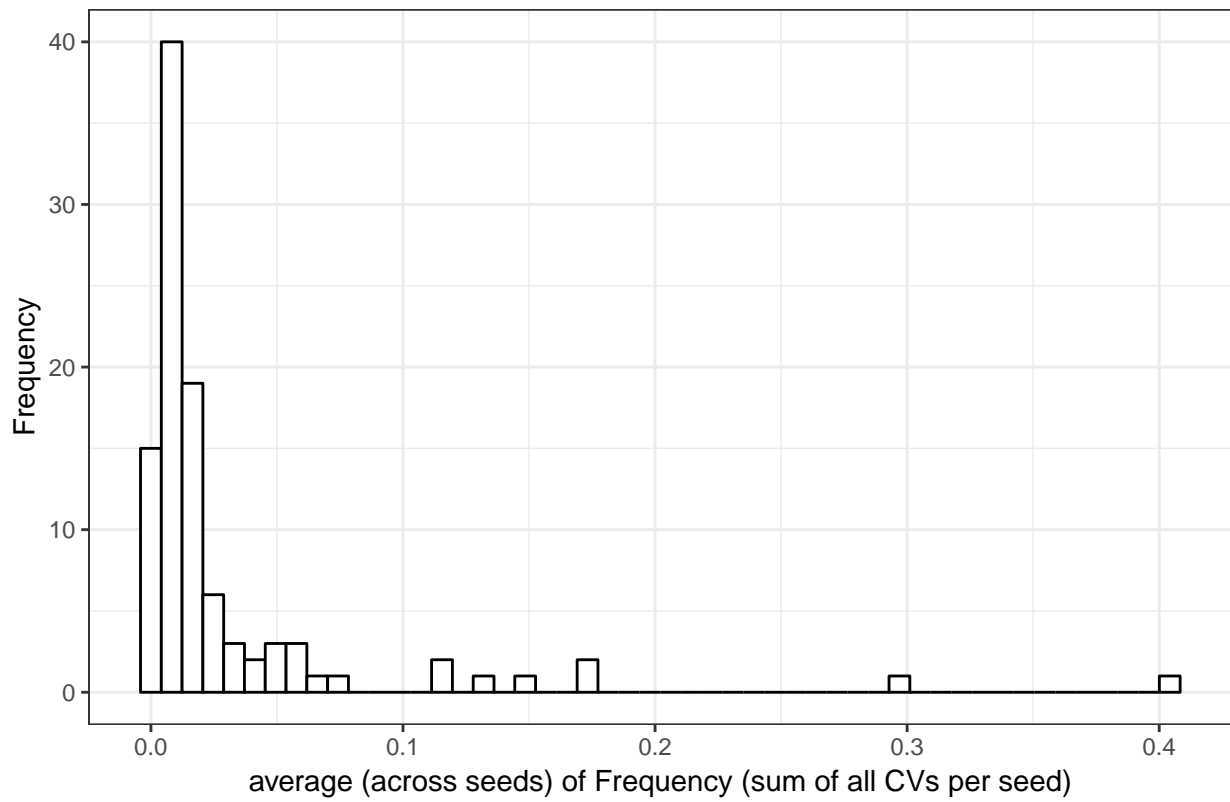
```
## Averaged pearson correlation across seeds# A tibble: 1 x 2
##   cor.avg cor.sdt
##   <dbl> <dbl>
## 1  0.227 0.0401
```

Correlation at seed = 1001

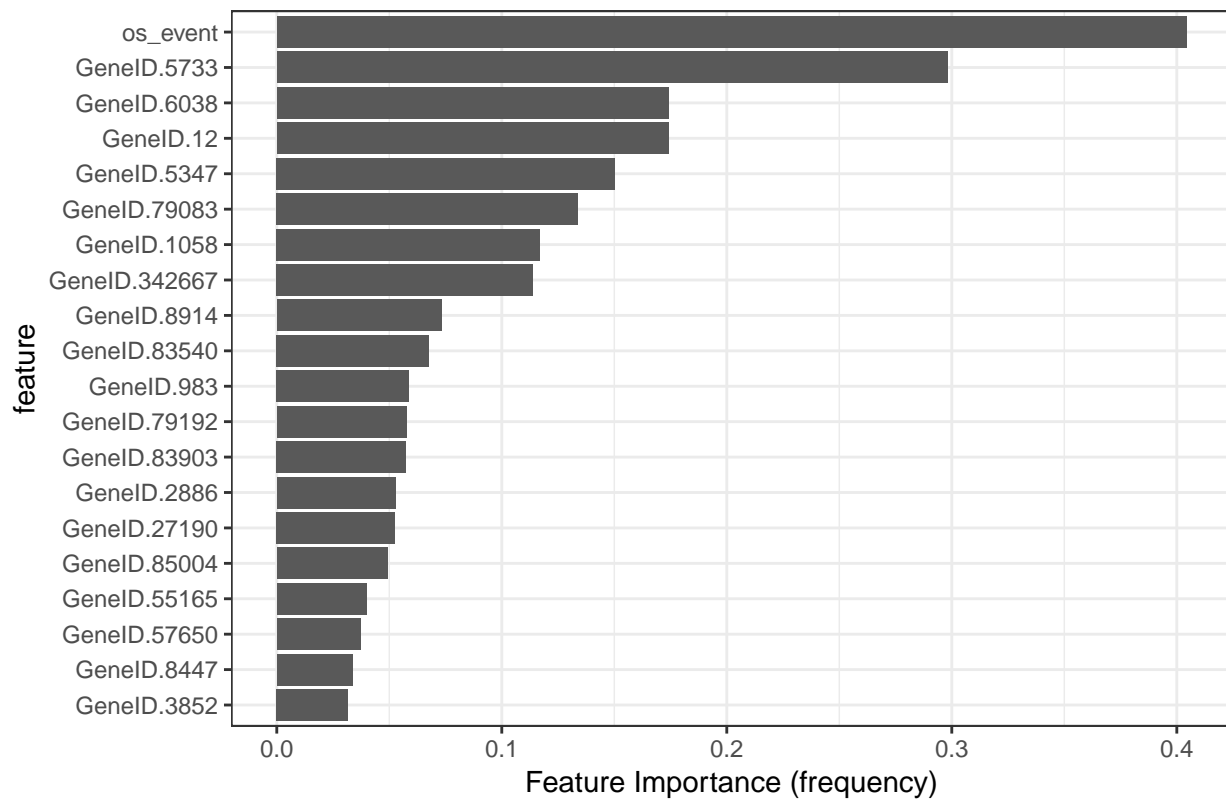


## 2. Important Features

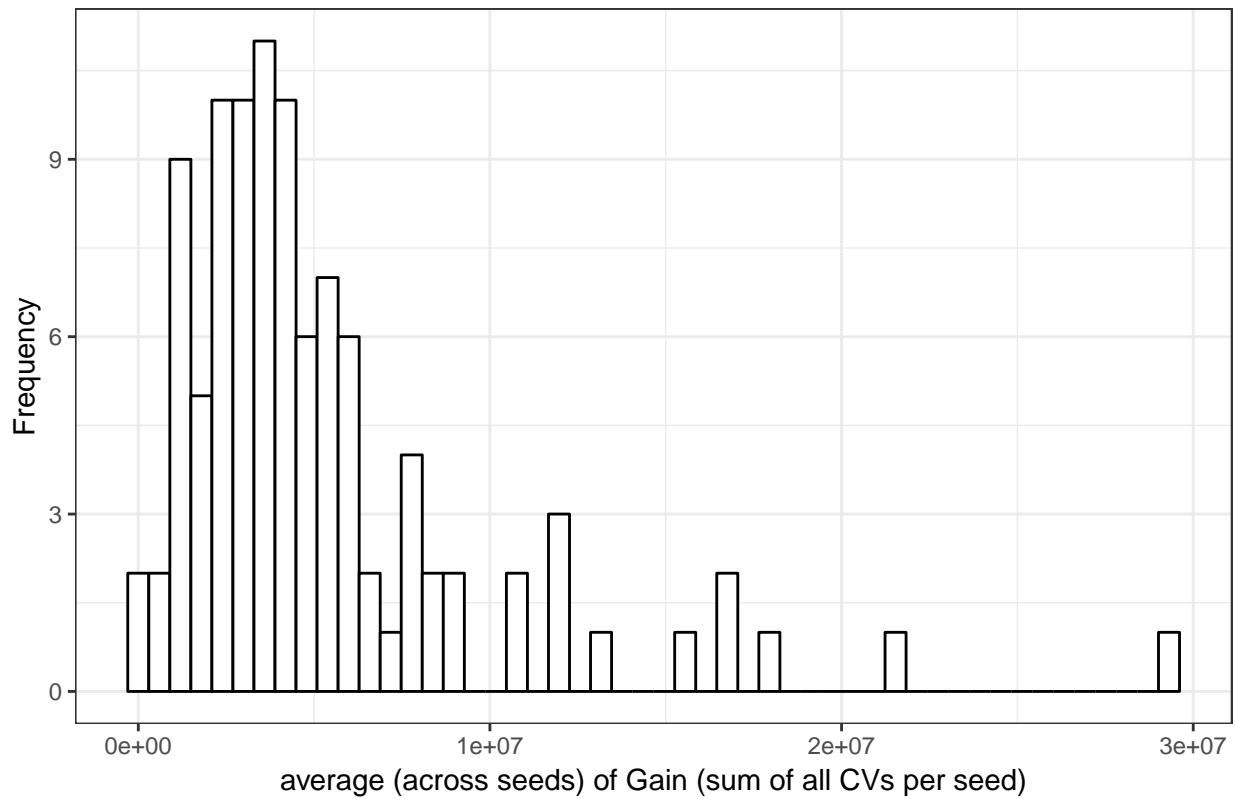
with 101 features based on Frequency



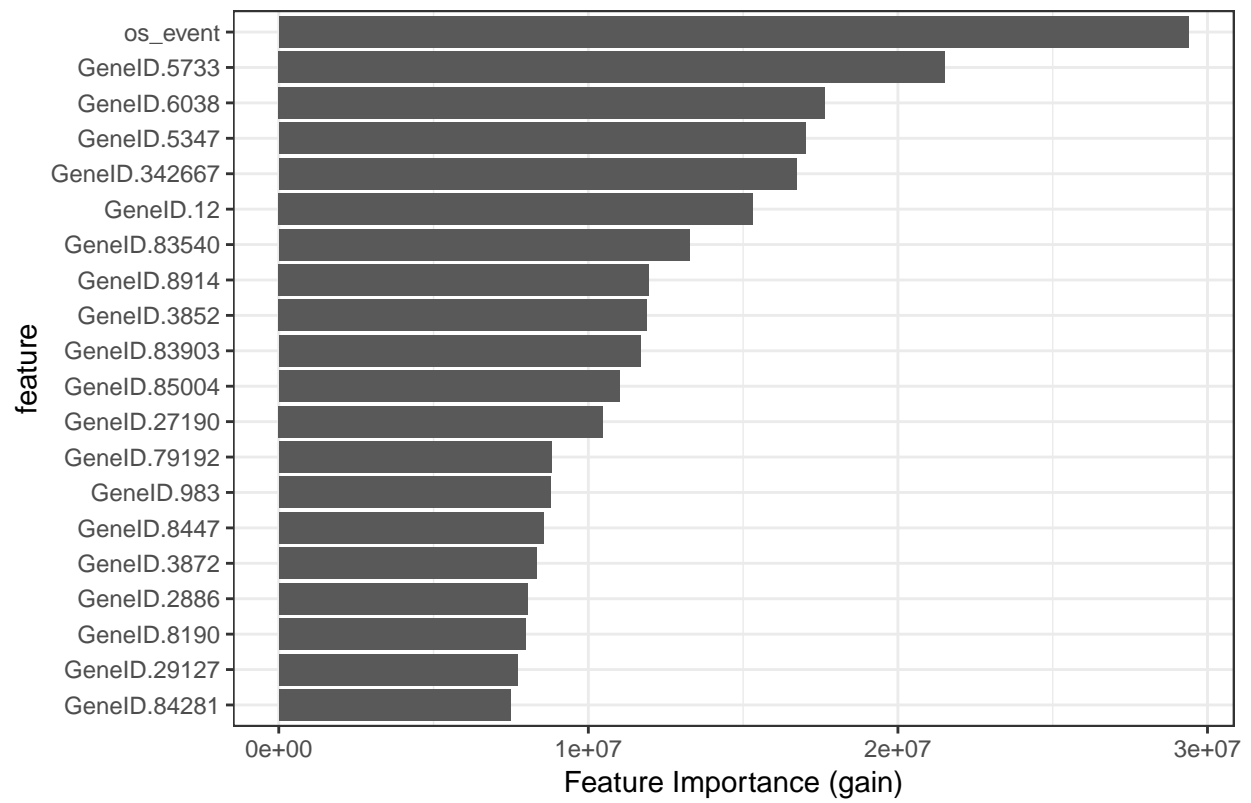
Top 20 features at 101 feature set based on Frequency

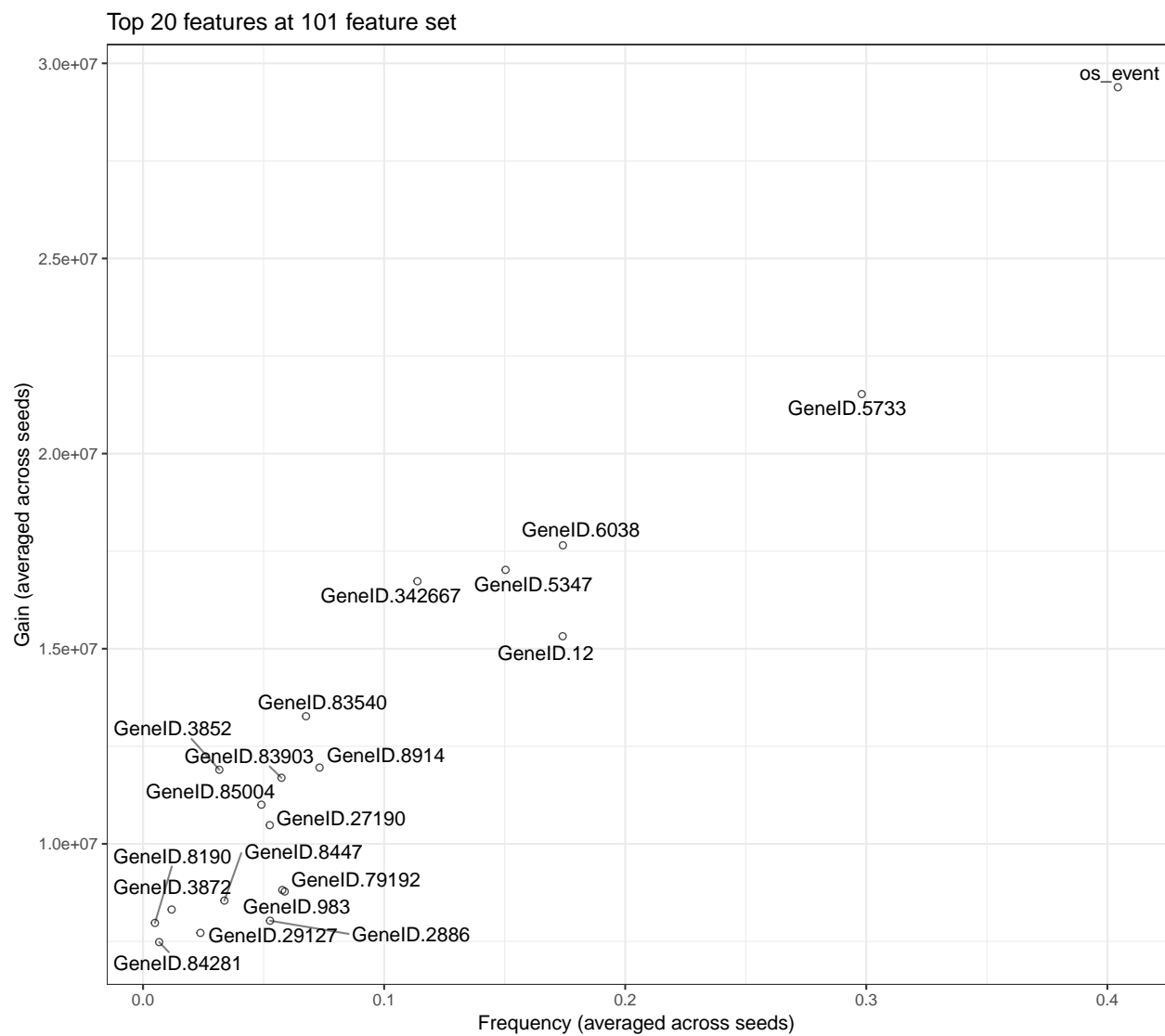


with 101 features based on Gain

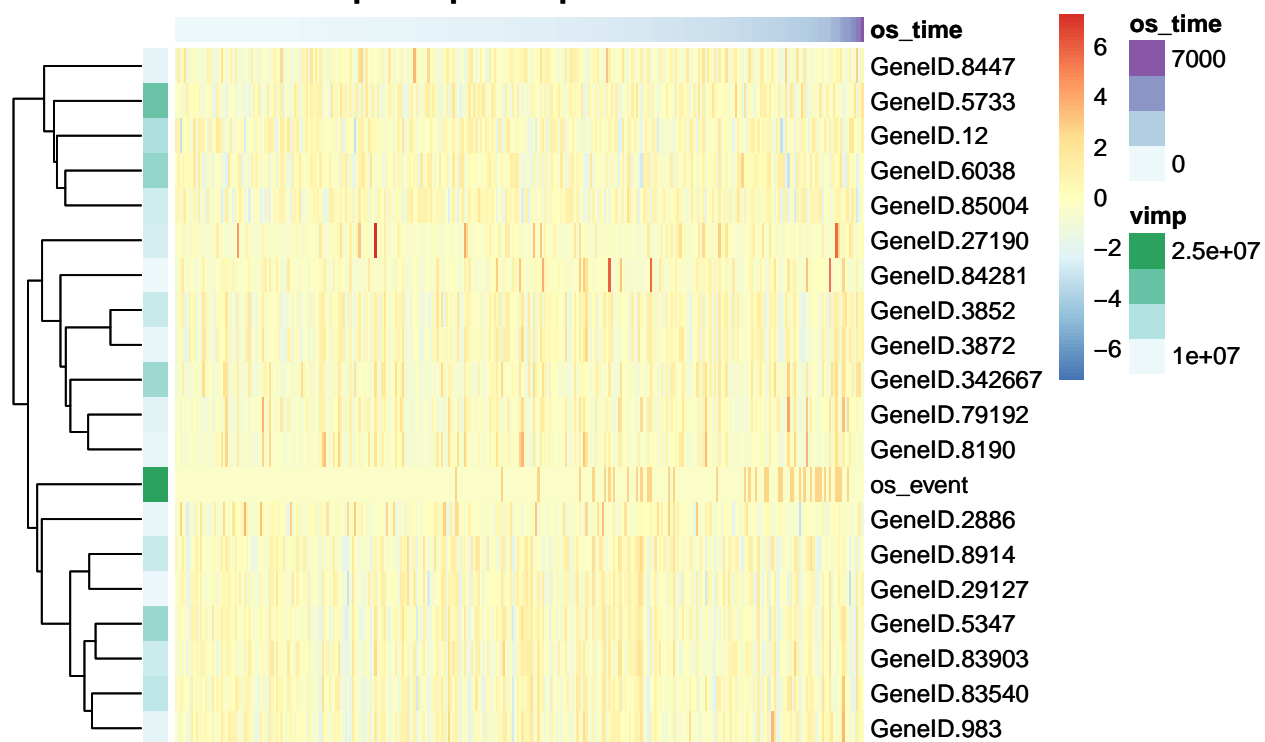


Top 20 features at 101 feature set based on Gain





# Heatmap of top 20 important features





### 3. Hyper-parameters

#### Hyperparameter Tuning in each CV

each box represents different seed (maximum 5 seeds are shown)

