

EIX :: Explain Interactions in XGBoost

The `EIX` package is the set of tools to structure mining of XGBoost and LightGBM models.

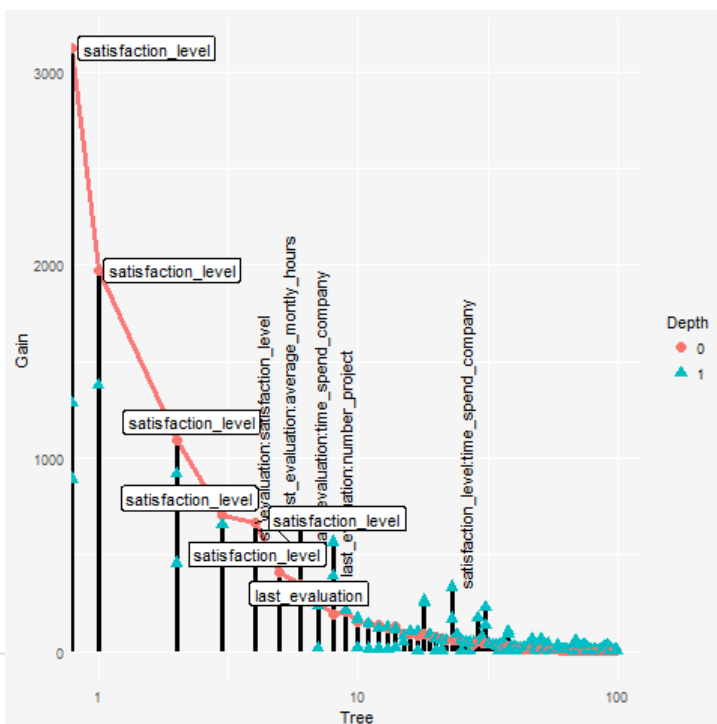
This package enables to find interactions in the model and also to measure variables' and interactions' importance. It includes the function of model visualization in such way that the most important variables and interactions are visible. It is also possible to analyze the prediction of a single observation taking into account the interactions. The last functionality was built using the `xgboostExplainer` and `breakDown` package.

```
library("Matrix")
library("data.table")
library("xgboost")
library("devtools")
install_github("ekarbowiak/EIX")
```

```
data<-data.table(HR_data, keep.rownames = F)
sparse_matrix <- Matrix::sparse.model.matrix(left~.-1, data = data)
label = data[,left]==1
xgb.train.data = xgb.DMatrix(sparse_matrix,label = label,missing = NA)
param <- list(objective = "binary:logistic", base_score = 0.5,max_depth=2)
xgb.model <-xgboost(param = param,data = xgb.train.data,nrounds = 50,verbose = FALSE)
```

Visualisation of model

```
lollipopPlot(xgb.model, sparse_matrix,
labels="topAll",log=TRUE)
```



Importance of variables and interactions

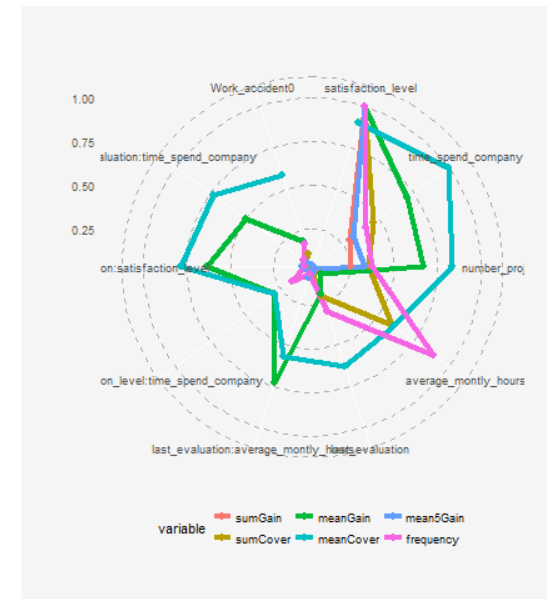
* opt="single" - object consists only single variables
* opt="interactions" - only interactions
* opt="mixed" - object shows importance both single variables and interactions. Default option.

```
importanceTable(xgb.model,
sparse_matrix, opt="mixed")
```

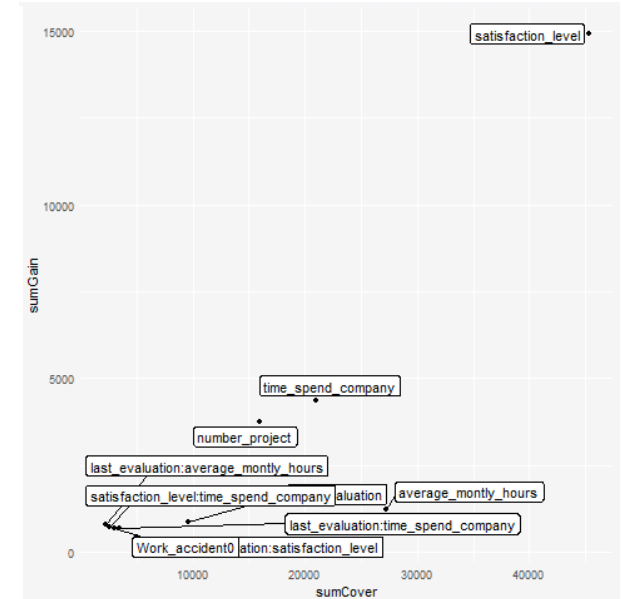
Available measures: sumGain, sumCover, mean5Gain (mean gain from 5 occurrences of given variable with the highest gain), meanGain, meanCover, frequency

For single variables additionally: counterRoots, meanDepth, weightedRoot (last two weighted by Gain of variable)

```
radarPlot(xgb.model,sparse_matrix,top=10,opt="mixed")
```



```
importancePlot(xgb.model, sparse_matrix,
xlab="sumCover", ylab="sumGain",
opt="mixed", top=10")
```

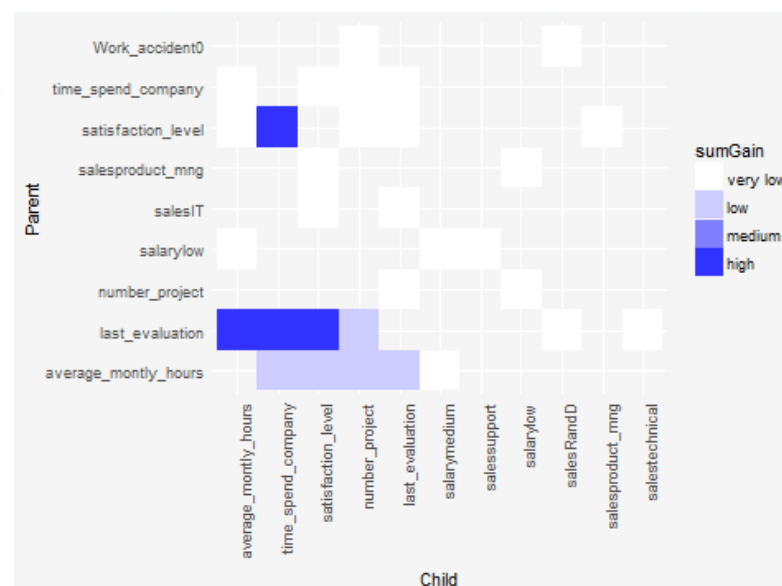


Interactions

```
interactionsPlot(xgb.model,sparse_matrix,
opt="interactions")
```

*opt="interactions" - these pairs of variables in which variable on the bottom (child) has higher gain than variable on the top (parent) . Default option.

*opt="pairs" - all pairs of variables, which occur in the model one above the other.



Prediction of single observation

```
id<-9
new_observation<-sparse_matrix[id,]
waterfallPlot(xgb.model, new_observation, opt="interactions")
```

