EIX:: Explain Interactions in XGBoost

MI

The **EIX** (**Explain Interactions in XGBoost**) package supports structure mining from XGBoost and LightGBM models.

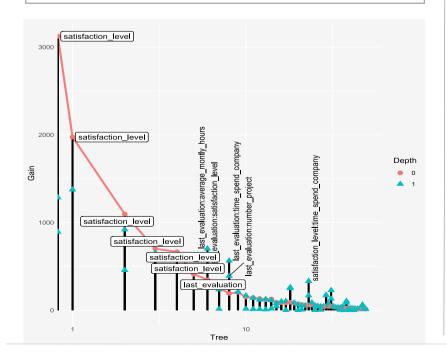
Key functionalities of this package cover: visualisation of tree-based ensembles models, identification of interactions, measuring of variable importance, measuring of interaction importance, explanation of single prediction with break down plots (based on **xgboostExplainer** and **breakDown** packages).

devtools :: install_github("ekarbowiak/EIX")
library("EIX")

Visualisation of model

The **lollipop plot** shows the model with labels of the most important variables and interactions.

lolli <- EIX_lollipop(xgb.model, sm)
plot(lolli, labels = "topAll", log_scale = TRUE)</pre>



Importance of variables and interactions

The **importance table** includes different measures of importance for variables and interactions. It is possible to visualise it in two ways using **plot**() function.

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

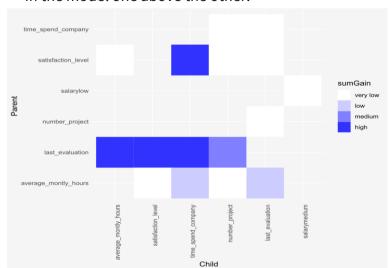
Available measures: sumGain, sumCover, mean5Gain (mean gain from 5 occurrences of given variable with the highest gain), meanGain, meanCover, frequency

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

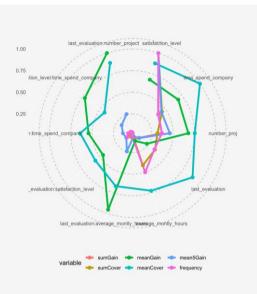
Interactions

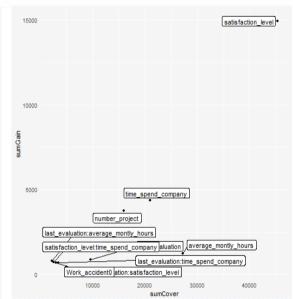
The **interactions table** makes the ranking of interactions in the model. Importance of interactions it can also visualise with **plot**() function.

- **option = "interactions"** these pairs of variables in which variable on the bottom (child) has higher gain than variable on the top (parent). Default option.
- **option = "pairs"** all pairs of variables, which occur in the model one above the other.



plot(imp, radar = FALSE, xmeasure = "sumCover", ymeasure = "sumGain", top = 10)





Explanation of a single prediction

The **waterfall plot** shows, which variables have influence on the prediction of a single observation.

