Interactive Exploration of Large Decision Tree Ensembles

Jan Forberg, Annett Mitschick, Martin Voigt*, and Raimund Dachselt

Interactive Media Lab, Technische Universität Dresden
* AI4BD Deutschland GmbH

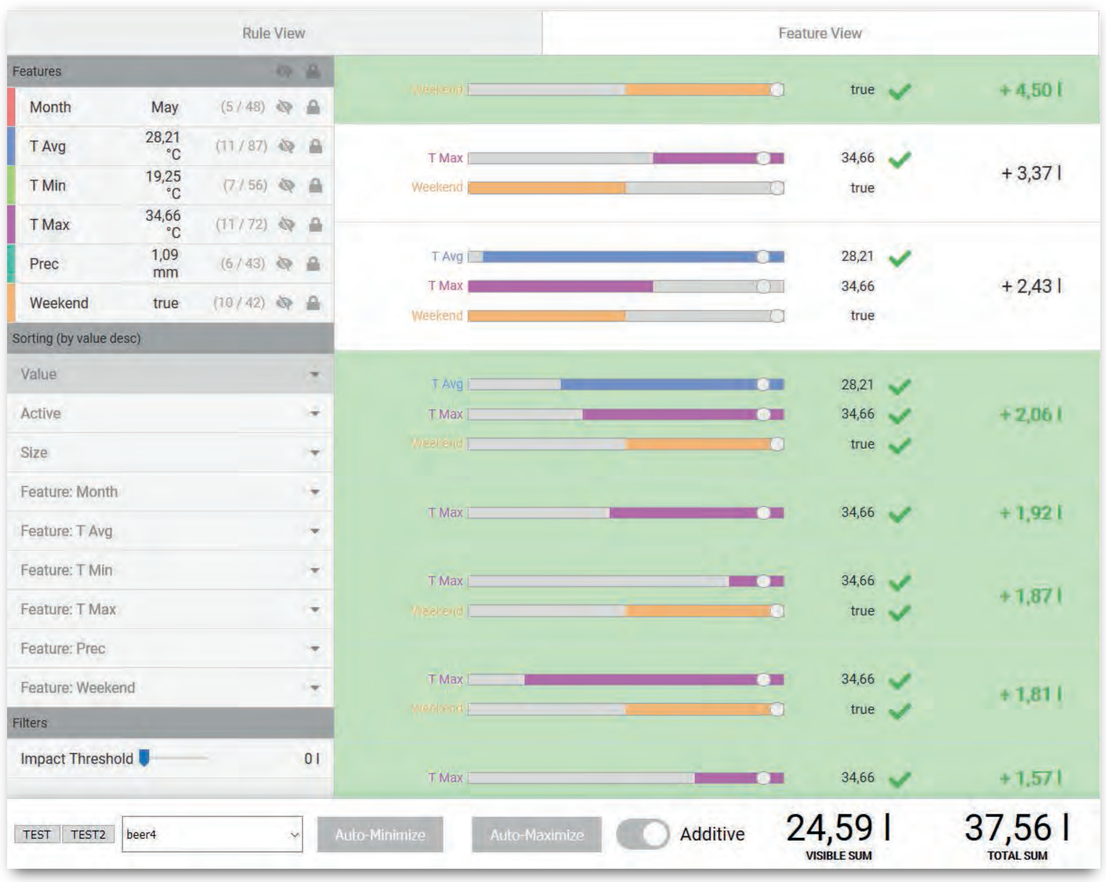
A visual exploration tool to make tree-ensemble based prediction models more interpretable and explicable

Visualization & Interaction Concept

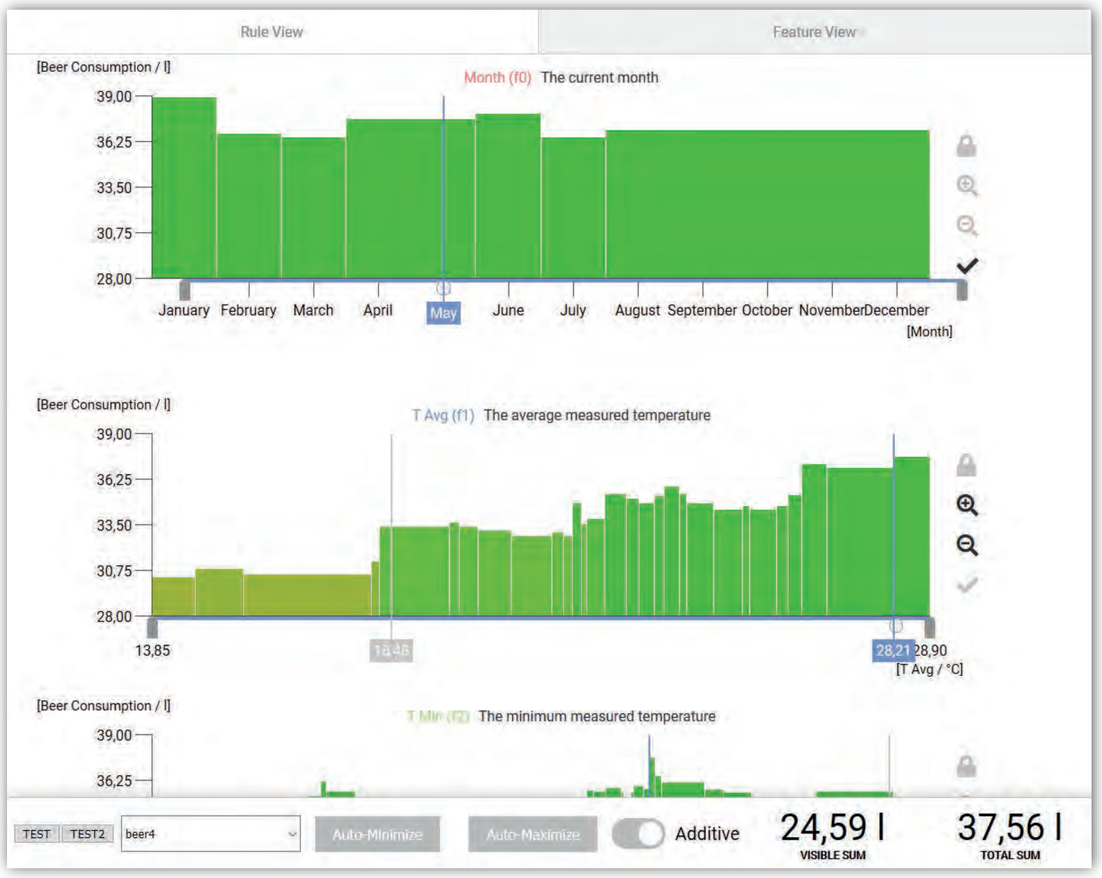
Creating a comprehensible visualization of a large forest consisting more than 100 decision trees is not straightforward. Consultants that implement and domain users that need to understand forcasting models (revenue, logistics, ...) are the target user groups we developed this concept for.
As decisions involving a specific feature are distributed throughout the ensemble, we propose a rule- and feature-centered visualization tool for model analysis.

The web-based prototype provides two interactive views:

Rule View



Feature View
with the „Dynamic Local Dependance Plot“



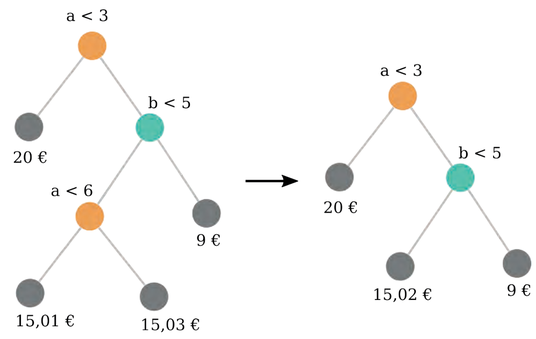
Contributions

- A visualization tool geared towards machine learning laypersons in order to make tree-based predictive models more comprehensible
- Dynamic Local Dependence Plots visualization
- An efficient rule extraction method to construct a substitute model for large tree-ensembles in the form of rule sets
- Preliminary results from an early think-aloud study with machine learning consultants

Rule Extraction

To convert a decision tree into a set of rules we interpret each distinct path from root to leaf as a rule, so that each split node denotes a rule condition.
Simplification methods are applied to reduce the size of the rule set as follows:

- ***Entropy Pruning:*** stop the traversal from root to leaf once a node with an entropy below a certain threshold has been found
- ***Impact Pruning:*** remove rules with values below a certain threshold
- ***Merge*** conditions targeting the same feature into a single condition.
- ***Merge*** two rules only differing in a single complementary condition into a single rule and a base prediction value.



Evaluation

We evaluated the preliminary prototype in a think-aloud study with five machine learning consultants.

They were asked to perform some exploration tasks with the visualization tool in a session of around 40-50 minutes.

- Especially the Feature View with the ***Dynamic Local Dependence Plots*** were very well-received
  → considered it to be very helpful for machine learning laypersons due to its simplicity
- top-down navigation from feature view (general overview) to rule view (detailed search) suggested as a reasonable extension of this work
- PSSUQ: best average scores were given for the statements “*I believe I could become productive quickly using this system*” (6,4) and “*The interface of this system was pleasant*” (6,4)

Evaluation of the rule extraction method:

By adjusting the thresholds, relatively small rule sets (less than 100 rules) could be created from forests consisting of around 100-150 trees while sustaining reasonable precision. We compared normalized mean deviation with the rule extraction methods [1] and [2].

Extensions

- A more central role of the feature view including means of navigation to the rule view
- A visual distinction of realistic and unrealistic data points
- Integration of Partial Dependence Plots to examine unusual local effects, providing a clear distinction between local and global behavior
- Exploration of the concepts on large displays
- Follow-up study with machine learning laypersons and a bigger group of participants

References

1 *Houtao Deng. 2018. Interpreting tree ensembles with inTrees. Int. Journal of Data Science and Analytics (2018), 1–11.*

2 *Satoshi Hara and Kohei Hayashi. 2018. Making Tree Ensembles Interpretable: A Bayesian Model Selection Approach. In Proc. of the 21st Int. Conf. on Artificial Intelligence and Statistics. 77–85. http://proceedings.mlr.press/v84/hara18a.html*

TECHNISCHE UNIVERSITÄT DRESDEN

INTERACTIVE MEDIA LAB DRESDEN

Contact information

Raimund Dachselt
raimund.dachselt@tu-dresden.de

Martin Voigt
martin.voigt@ai4bd.com

ai4bd