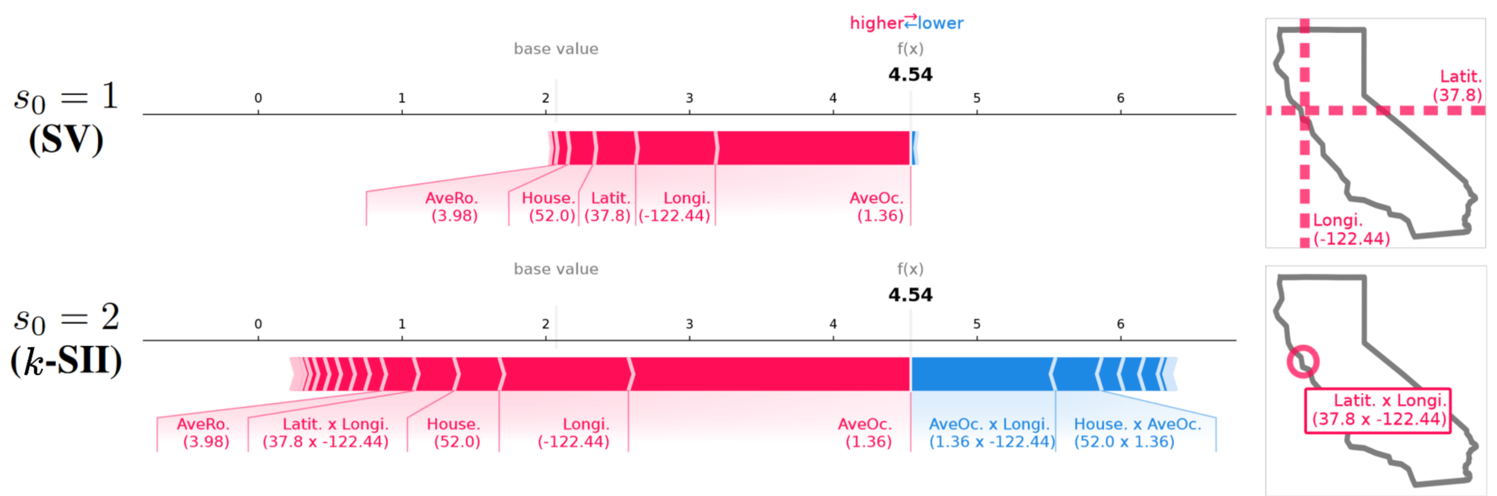
Beyond TreeSHAP: Efficient Computation of Any-Order Shapley Interactions for Tree Ensembles

Maximilian Muschalik[1,*], Fabian Fumagalli[2,*], Barbara Hammer[2] Eyke Hüllermeier[1]

Association for the Advancement of Artificial Intelligence

Interaction Example: Explaining Property Prices

Potential Question: "Does the *location* of my property affect its price?"



TreeSHAP: Application of Game Theory to Feature Attributions for Tree Ensembles

Features

$N = \{1, \dots, n\}$

Decision Tree

$f : \mathcal{X} \to \mathbb{R}$

Local Explanation

for $x \in \mathcal{X}$

Extended Decision Tree with Missing Features $f : \mathcal{X} \times \mathcal{P}(N) \to \mathbb{R}$

$$f(T) := f(x, T) := \sum_{v \in \text{leaf nodes}} (\text{leaf prediction}) \cdot (\text{ratio reaching } v \text{ given } T\ )$$

Interpretation: *average prediction* given only features $T \subseteq N$

Weighted predictions based on split ratios for features in $N \setminus T$ **[1,2]**

Extended Decision Tree is a Set Function and Cooperative Game Theory can be applied

Shapley Value [3]

Assigns axiomatic fair individual contributions

The Shapley Value is the average (T-weighted) **marginal contribution**: $f(T \cup \{i\}) - f(T)$

Shapley Interaction Index (SII) [4]

Assigns axiomatic interactions to groups complying with Shapley principles

The SII is the average (T-weighted) **discrete derivative**: $\sum_{L \subseteq S} (-1)^{|S|-|L|} f(T \cup L)$

Constructing Interpretable Interaction Indices from SII

k-Shapley Values (k-SII) [5]

Assigns **contributions** to groups of size $\leq k$ by **aggregating SII** of size $\leq k$,

which yields an

Interpretable k-order Interaction Index

k-SII Interpretation

(average) change in the model's prediction

due to the **joint information** provided by

a group of features of size $\leq k$, which

cannot be attributed to any subgroup

References

[1] Yu, P., Bifet, A., Read, J., Xu, C., (2022). Linear tree shap. Advances in Neural Information Processing Systems 35.
[2] Lundberg et al., (2020). From local explanations to global understanding with explainable AI for trees. In Nat. Mach. Intell. pp. 56–67.
[3] Shapley, L. S. (1953). A value for n-person games. In Contributions to the Theory of Games (AM-28), Volume II, pages 307–318. Princeton University Press.
[4] Grabisch, M. and Roubens, M. (1999). An axiomatic approach to the concept of interaction among players in cooperative games. International Journal of Game Theory, 28(4):547–565
[5] Bordt, S. and von Luxburg, U. (2023). From shapley values to generalized additive models and back. In AISTATS'23, pp. 709–745.
[6] Fujimoto, K., Kojadinovic, I., and Marichal, J. (2006). Axiomatic characterizations of probabilistic and cardinal-probabilistic interaction indices. Games Econ. Behav., 55(1):72–99.
[7] Fumagalli, F., Muschalik, M., Kolpaczki, P., Hüllermeier, E., Hammer, B. (2023). SHAP-IQ: Unified Approximation of any-order Shapley Interactions Advances in Neural Information Processing Systems 36.

LMU Ludwig-Maximilians-Universität München
Universität Bielefeld
Funded by DFG Deutsche Forschungsgemeinschaft German Research Foundation

1) LMU Munich, Munich, Germany 2) Bielefeld University, Bielefeld, Germany
* equal contribution

Contribution

TLDR: We propose **TreeSHAP-IQ**: An **efficient** algorithm for computing **any-order Shapley Interactions** [3-7] for tree ensembles similar to **Linear TreeSHAP** [1,2].

TreeSHAP-IQ: Efficient Computation of Any-Order Shapley Interactions

TreeSHAP [2]

reduces the Shapley Value complexity from **exponential** to **quadratic** (in #nodes) using a **recursive** tree traversal

Linear TreeSHAP [1]

improves TreeSHAP to **linear** (in #nodes) using **polynomial arithmetic**

TreeSHAP-IQ

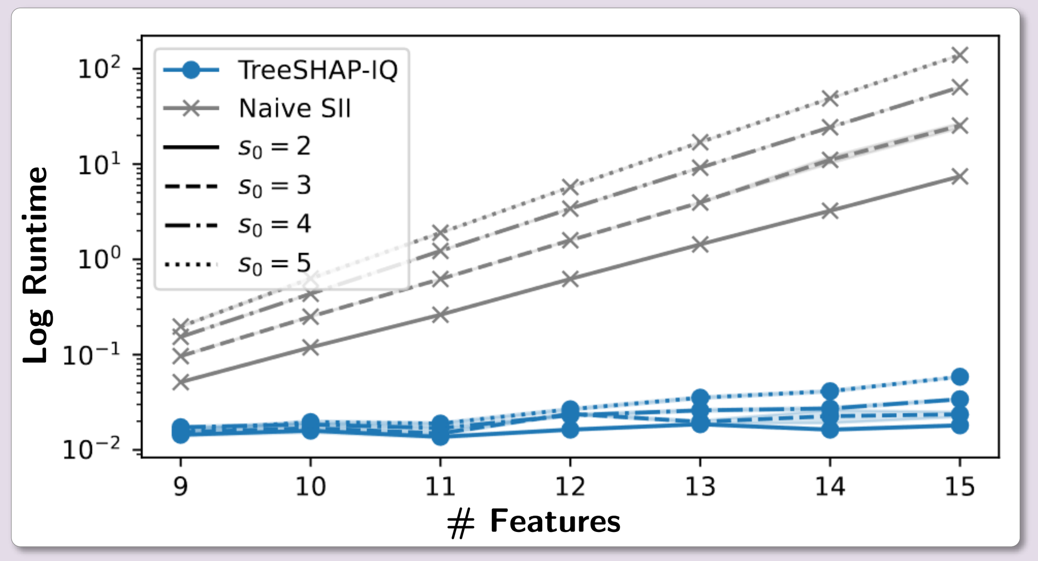
introduces **novel polynomials** to compute **any-order SII** as well as general **interaction indices**, including **Banzhaf Interactions** in **linear time** (in #nodes per interaction) yielding a **k-order additive local explanation**

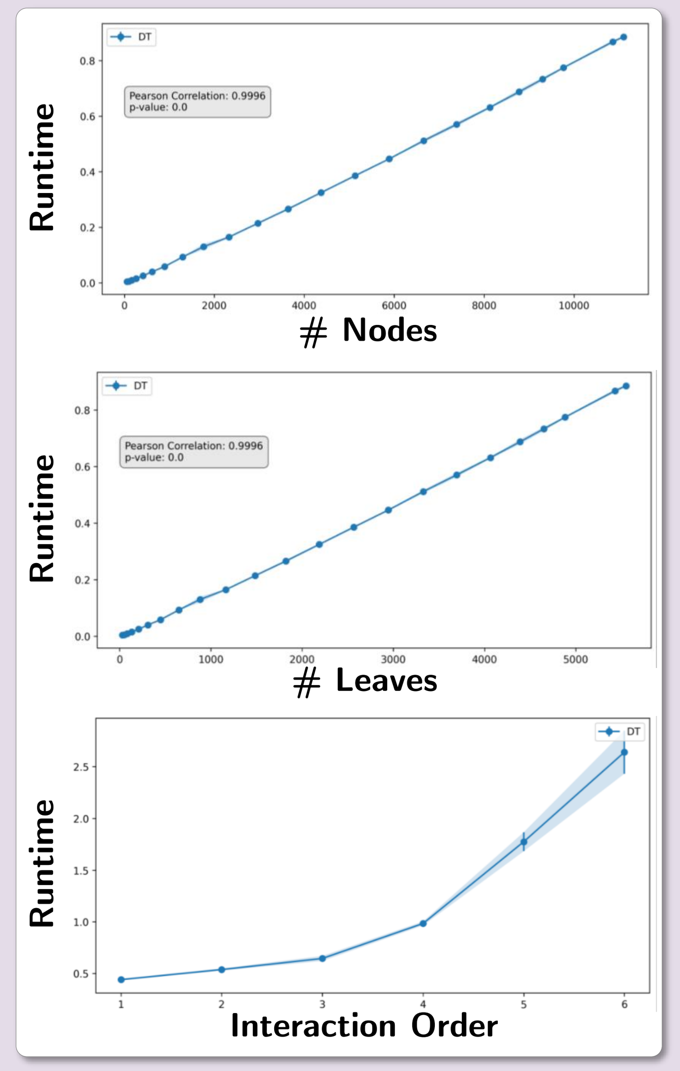
Complexity and Runtime of TreeSHAP-IQ

Complexity	TreeSHAP-IQ	Linear TreeSHAP
Comp.:	$\mathcal{O}\left(m \cdot \ell_T \cdot d_{\max} \cdot \binom{n-1}{s-1}\right)$	$\mathcal{O}(m \cdot \ell_T \cdot d_{\max})$
Storage:	$\mathcal{O}\left(d_{\max}^2 \cdot \binom{n}{s}\right)$	$\mathcal{O}(d_{\max}^2 + n)$

n: number of features $\quad \ell_T$: number of leaves $\quad m$: explanation points

s: interaction order $\quad d_{\max}$: max tree depth

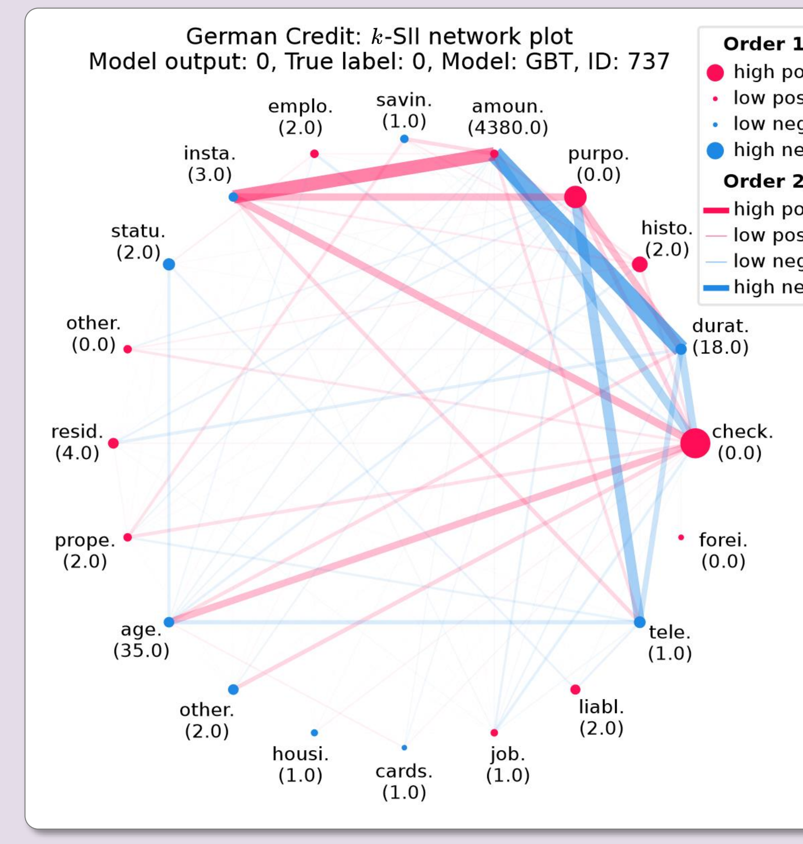




Datasets	# Instances	# Features	Target	Speed-Up
Credit	1 000	20	$\{0, 1\}$	$\sim 10^4$
Bank	45 211	16	$\{0, 1\}$	$\sim 10^3$
Adult	45 222	14	$\{0, 1\}$	$\sim 10^3$
Bike	17 379	12	$\mathbb{R}$	$\sim 10^1$
COMPAS	6 172	11	$\{0, 1\}$	$\sim 10^2$
Titanic	891	9	$\{0, 1\}$	$\sim 10^1$
California	20 640	8	$\mathbb{R}$	~ 1

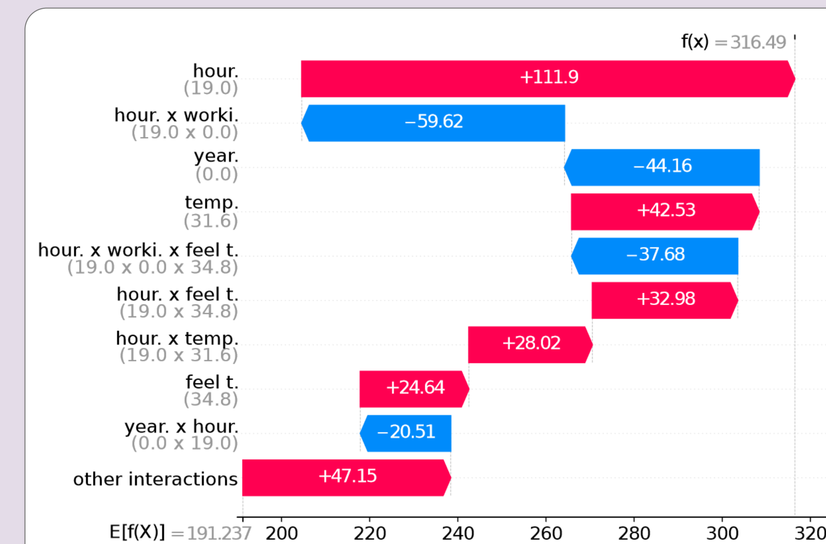
Features ↑ Speed-Up

Visualize TreeSHAP-IQ Interactions



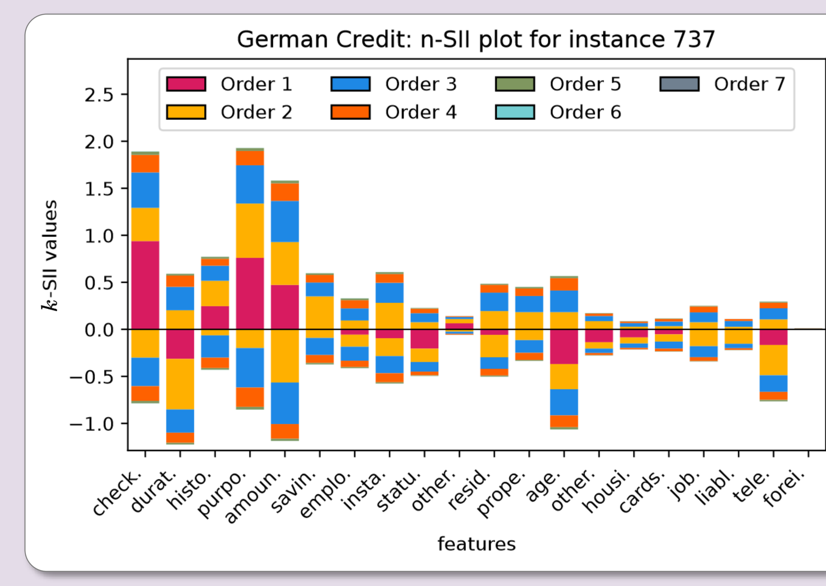
Network Plot

- describes 1st order (vertices) and 2nd order (edges) contribution towards the prediction
- **here**: high positive and negative interaction influence the prediction



Waterfall Plot

- breaks down the contribution of features and interactions
- **here**: 2nd and 3rd order interactions most important for prediction



k-SII Stacked Bar Plot

- illustrates how much interaction is present for a single instance
- **here**: only negligible interactions after order 4

Open Source Implementation

TreeSHAP-IQ is available for python

```
pip install shapiq
```

- **compute** general Shapley interactions for SOTA tree-based models
- **plot interactions**