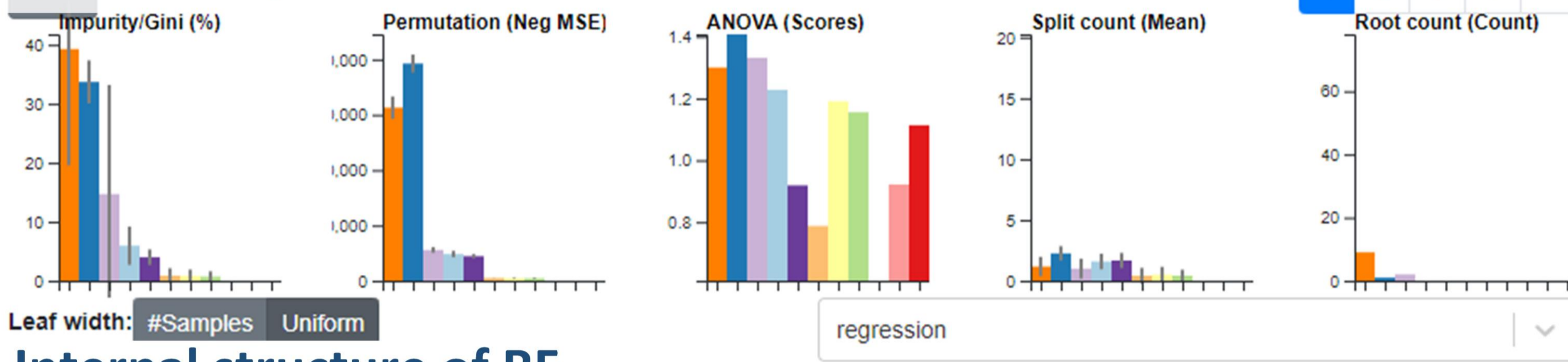


Visualizing Prediction Provenance in Regression Random Forests

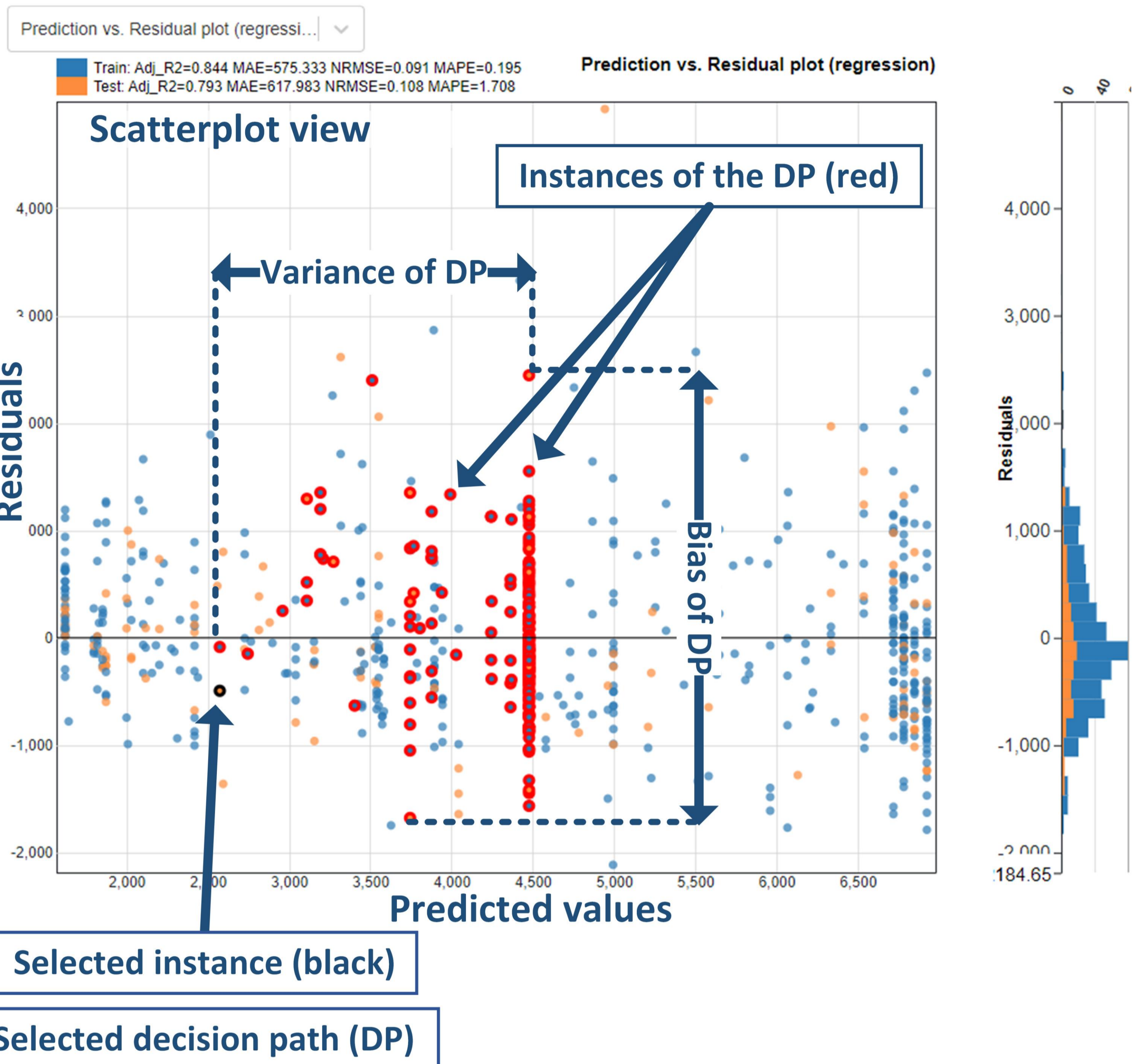
Table of RF models

Name	max_depth	max_leaf_nodes	n_estimators	adj_r2_test	mae_test	nrmse_test	mape_test
bike_sharing_day_RF0	50	10	12	0.793	617.983	0.108	1.708
bike_sharing_day_RF1	50	10	30	0.795	620.506	0.107	1.7
bike_sharing_day_RF2	50	10	49	0.775	640.793	0.112	1.898
bike_sharing_day_RF3	50	10	100	0.781	633.882	0.111	1.835

Feature importance



Internal structure of RF



Problem: provenance of random forest (RF) predictions is complex to apprehend by end users who need to build trust in the model

Solution: Multiple coordinated views supporting prediction provenance analysis

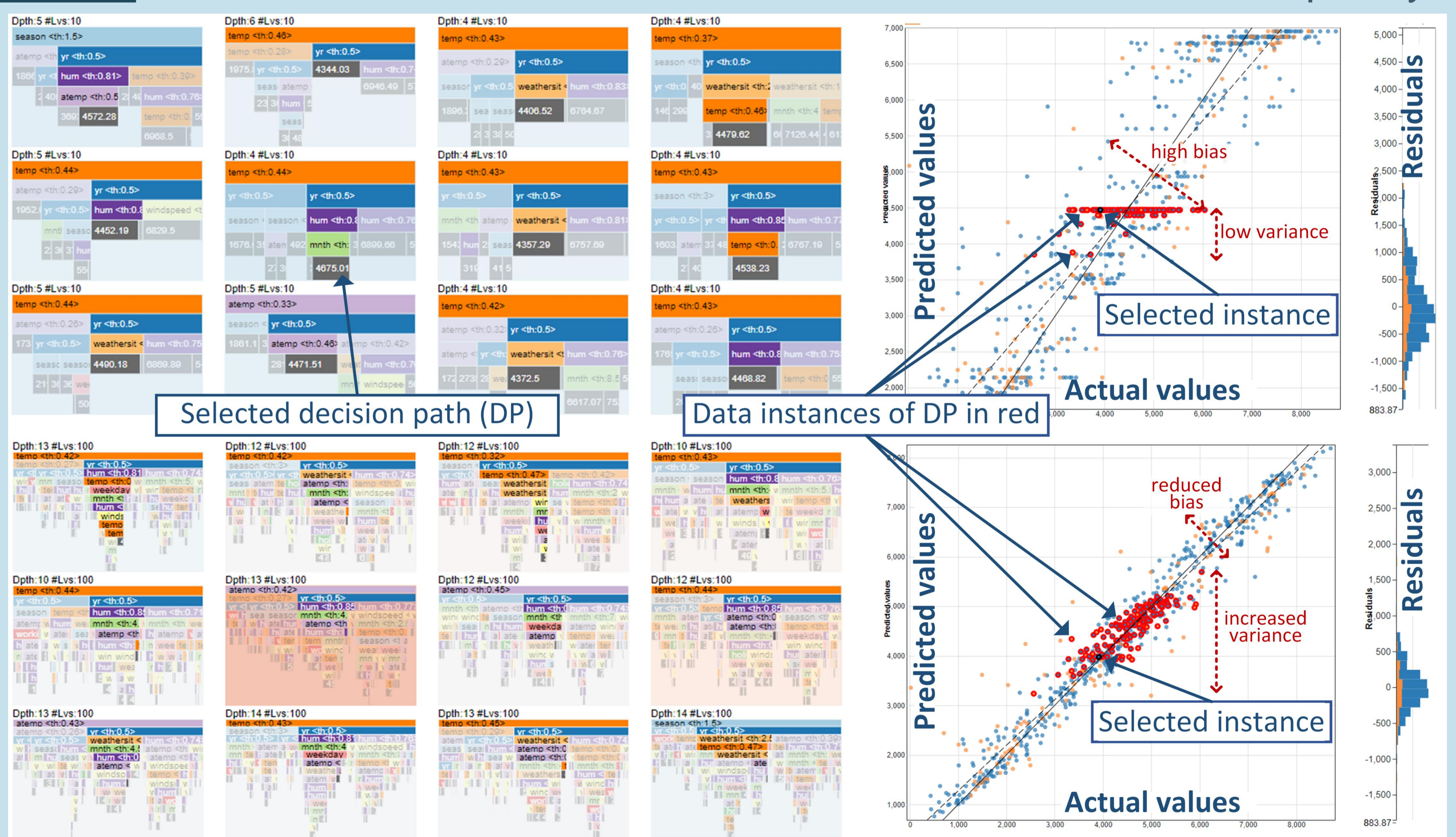
- Level 1: inspect data instance
- Level 2: inspect decision path
- Level 3: inspect individual decision tree
- Level 4: inspect global random forest

Goals related to **trust levels (TL)** [1]

- **G1** - Rules / criteria used by the model
Understanding and explanation (TL3)
- **G2** - Data used to learn the model or form a decision path (DP)
Diagnosis (TL3)
- **G3** - Degree of uncertainty for a prediction
Performance, model bias and variance (TL4)

Variation of hyper parameters: when max nb. of leaves \uparrow , bias \downarrow but variance and RF complexity \uparrow

- max. nb. leaves: 10
- nb trees: 12
- RF model performance: Adjusted $R^2 = 0.793$
- max. nb. leaves: 100
- nb trees: 12
- RF model performance: Adjusted $R^2 = 0.879$



References

[1] CHATZIMPARMPAS A., MARTINS R. M., JUSUFI I., KUCHER K., ROSSI F., KERREN A.: The State of the Art in Enhancing Trust in Machine Learning Models with the Use of Visualizations. Computer Graphics Forum 39, 3 (2020), 713–756