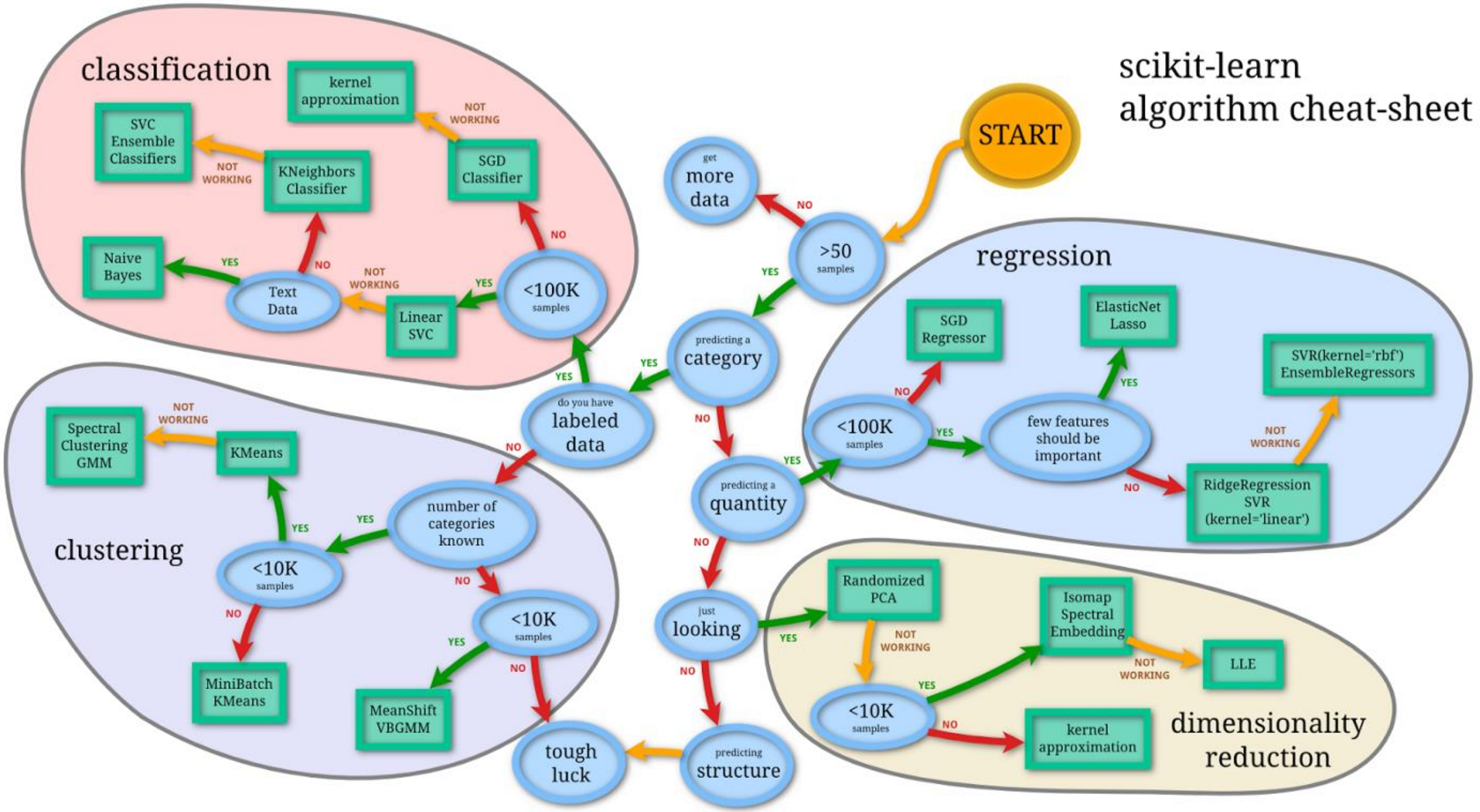
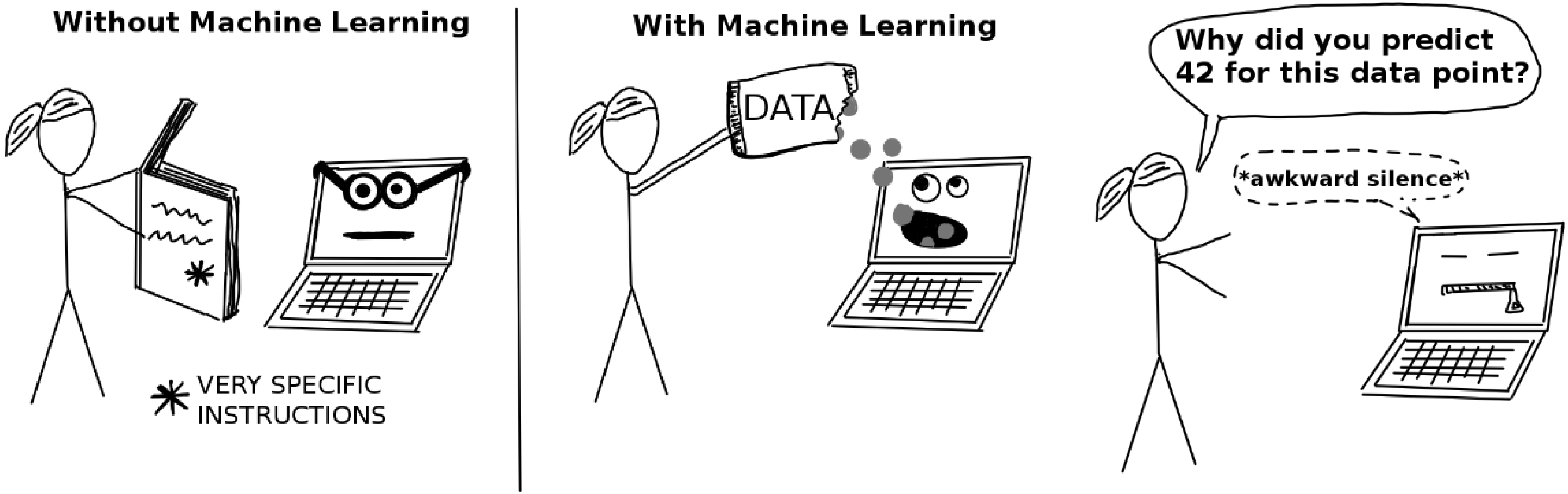
Bagging, boosting, and stacking

Chemical Engineering
& Materials Science

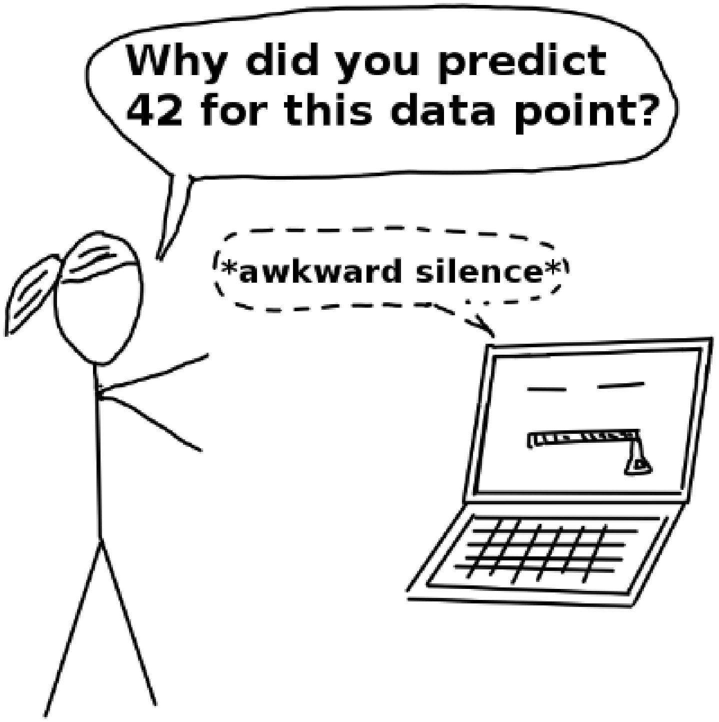
Which model to use??



General approaches to interpretable ML



Why do we care about interpretability?



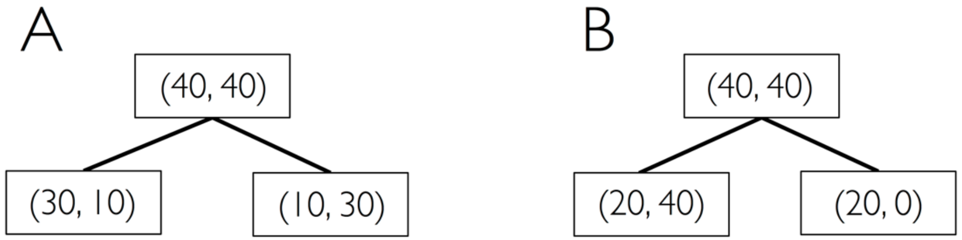
Taxonomy of interpretability

Intrinsic interpretability in linear regression

$$y \approx \hat{y} = f(X) = w_0 x_0 + w_1 x_1 + \cdots$$

$$w = w: \min_w \left|\left|\hat{y} - y\right|\right|_2^2$$

Intrinsic interpretation of decision trees



From model-specific to model-agnostic methods

One global method: permutation importances

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici. "All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously." http://arxiv.org/abs/1801.01489 (2018)

One global method: permutation importances

Compute on training data or validation data?

Fisher, Aaron, Cynthia Rudin, and Francesca Dominici. "All models are wrong, but many are useful: Learning a variable's importance by studying an entire class of prediction models simultaneously." http://arxiv.org/abs/1801.01489 (2018)

One global method: permutation importances

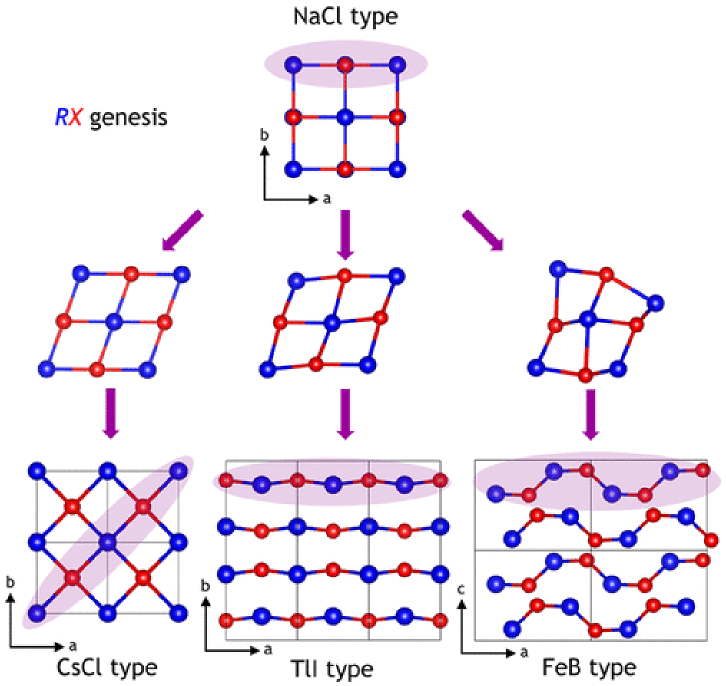
CEMS
Chemical Engineering
& Materials Science

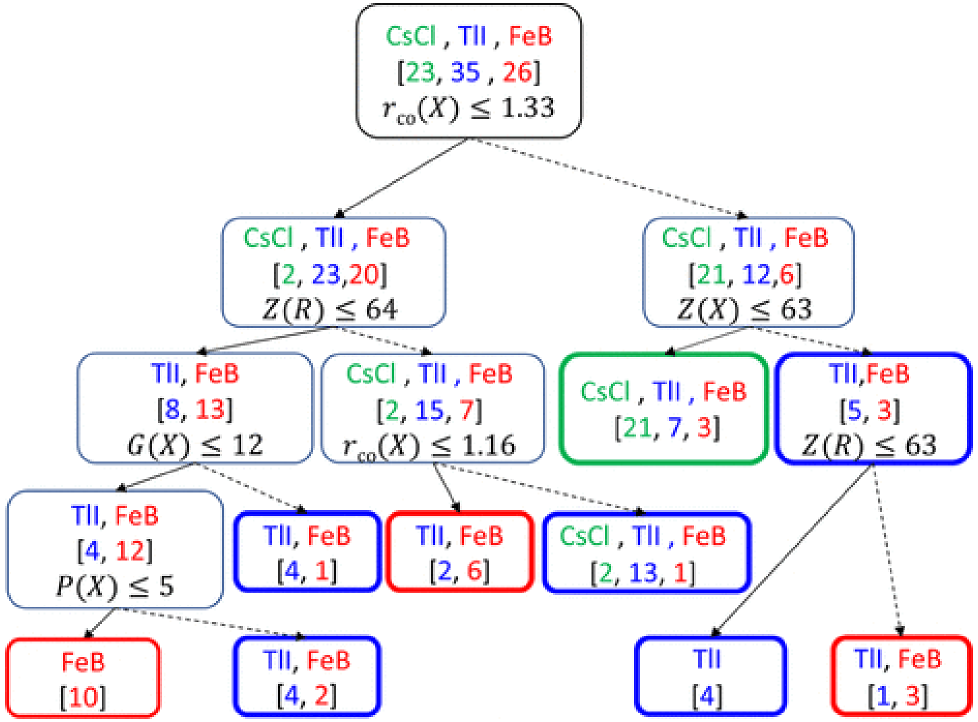
CHEMISTRY OF MATERIALS
pubs.acs.org/cm Article

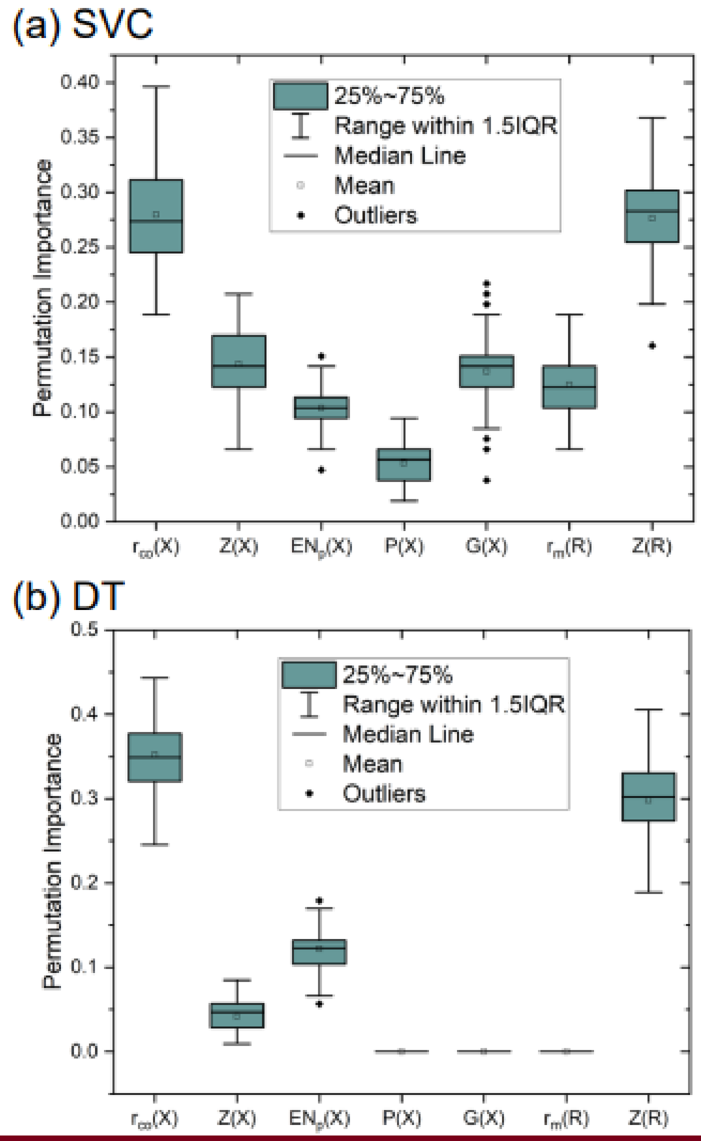
Revealing Hidden Patterns through Chemical Intuition and Interpretable Machine Learning: A Case Study of Binary Rare-Earth Intermetallics *RX*

Volodymyr Gvozdetskyi,* Balaranjan Selvaratnam, Anton O. Oliynyk, and Arthur Mar*

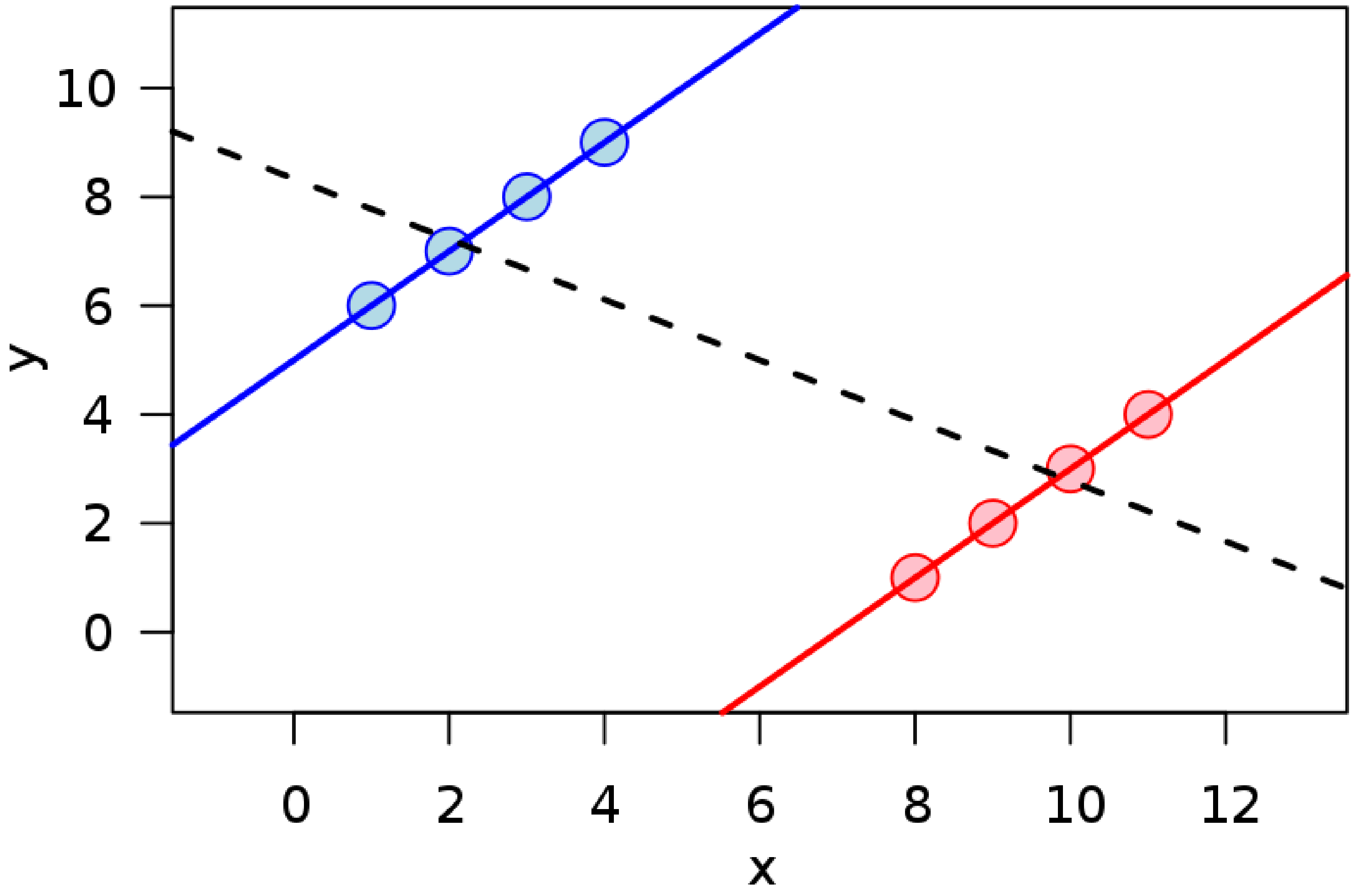
Cite This: Chem. Mater. 2023, 35, 879–890 Read Online





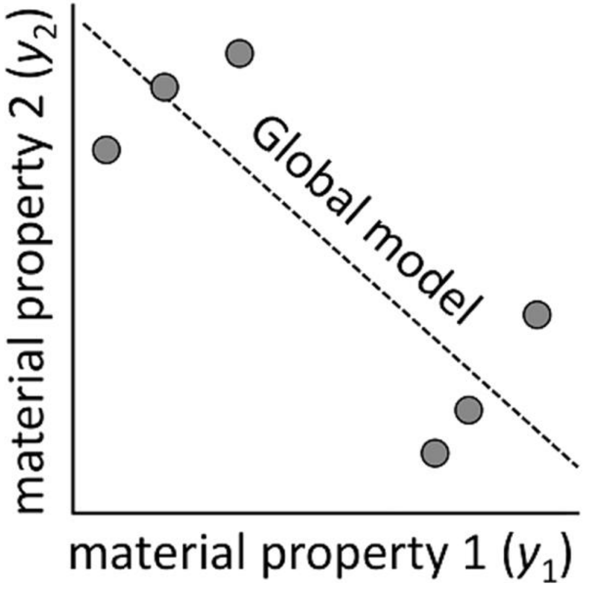


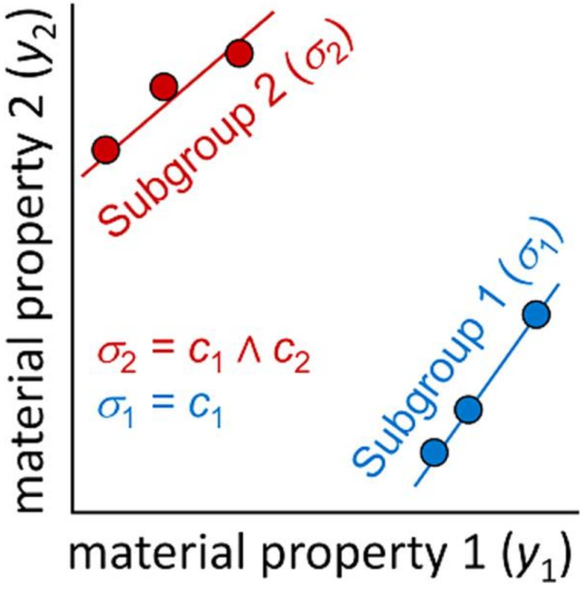
From global to local methods

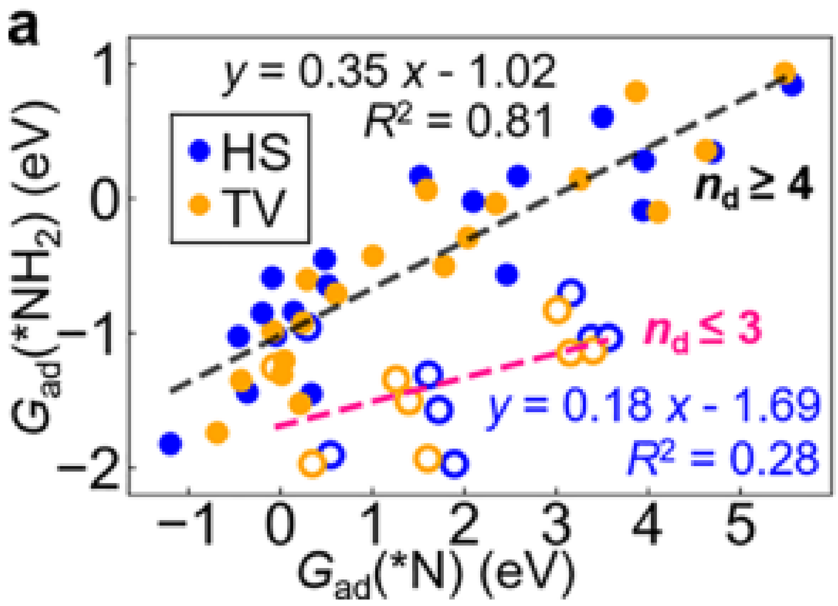


https://en.wikipedia.org/wiki/Simpson's_paradox

Our data may not be globally interpretable







 Goldsmith et al., New Journal of Physics, 2017, 19, 013031 | Li et al., ACS Catalysis, 2021, 11, 7906

Local version of surrogate: LIME

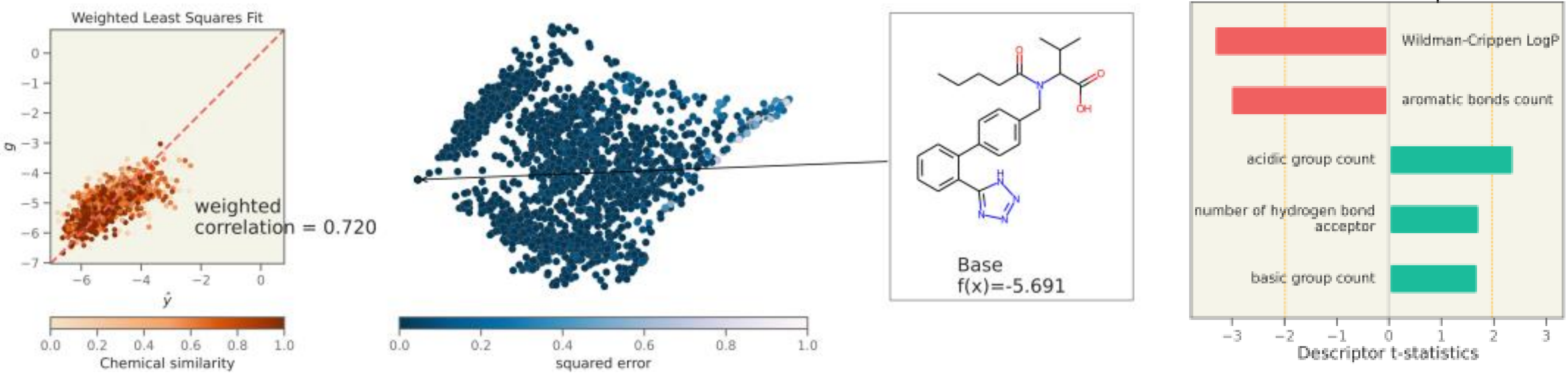
Local version of surrogate: LIME

Explaining molecular properties with natural language

A Preprint

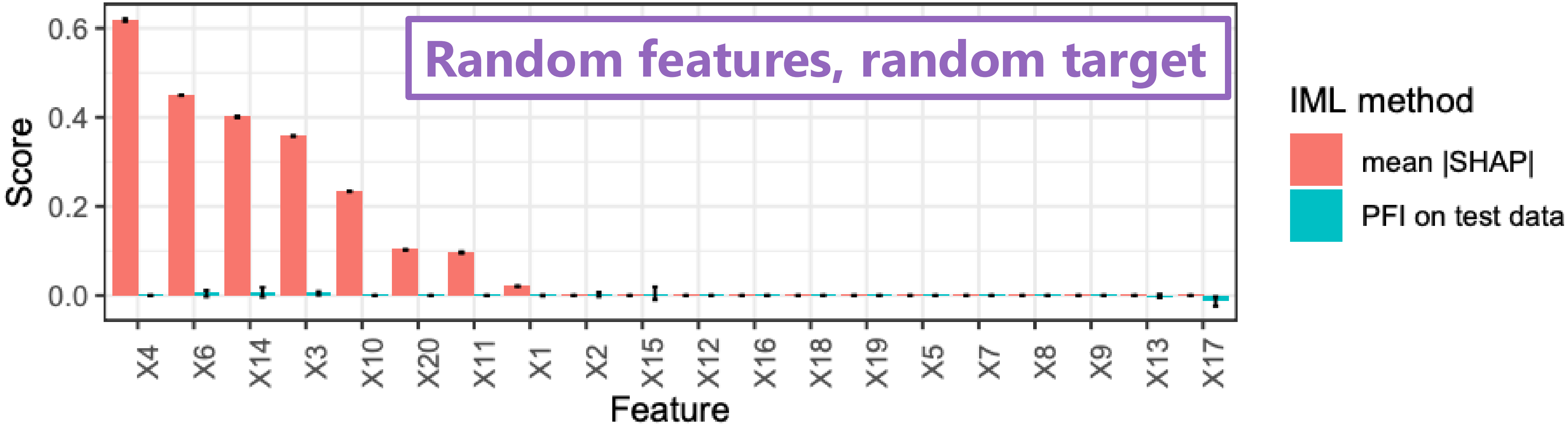
Heta A. Gandhi
Department of Chemical Engineering
University of Rochester
Rochester, NY, 14627
hgandhi@ur.rochester.edu

Andrew D. White*
Department of Chemical Engineering
University of Rochester
Rochester, NY, 14627
andrew.white@rochester.edu



Pitfalls of interpretable ML

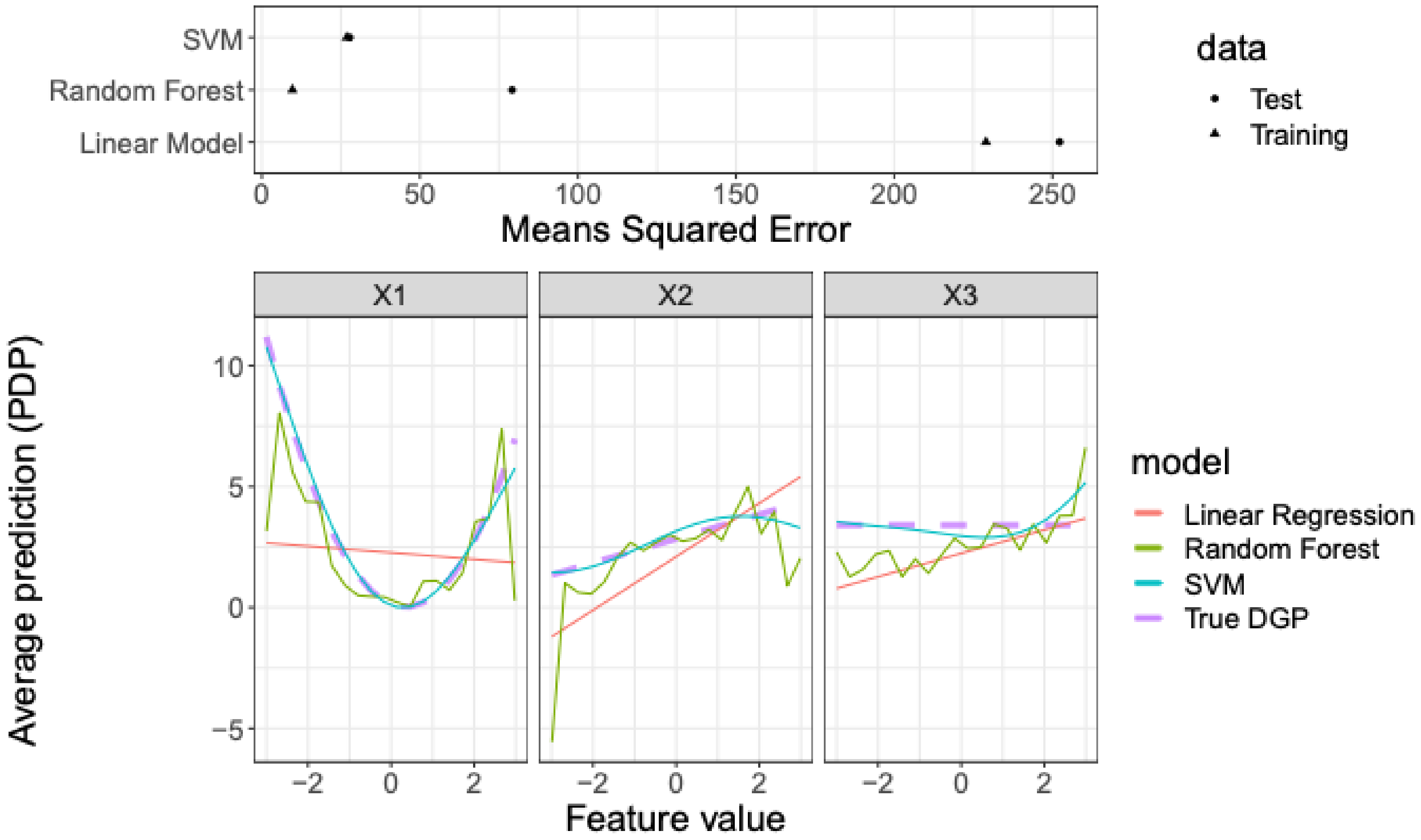
1. Assuming one method will always work



Molnar et al., 2020, arXiv:2007.04131

Pitfalls of interpretable ML

2. Bad model generalization



Molnar et al., 2020, arXiv:2007.04131

Pitfalls of interpretable ML

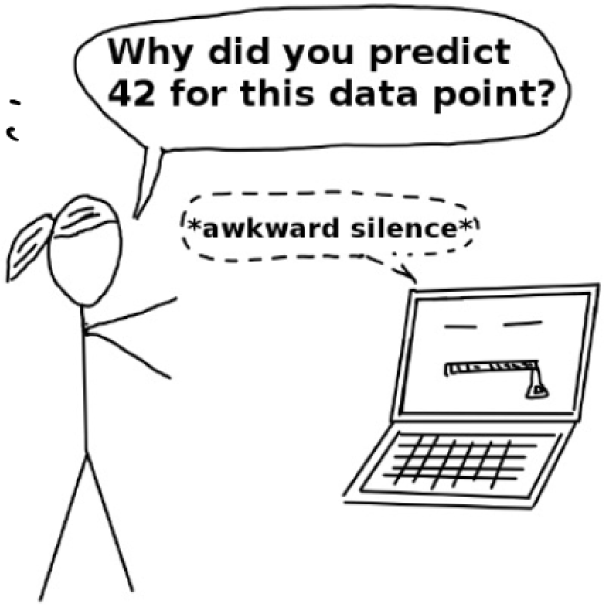
3. Unnecessary complexity

Recall why we care about interpretability:

Reliability → small change in x_i shouldn't lead to a large change in $\hat{y}_i$

Causality → as we change x_i, can we anticipate change in $\hat{y}_i$

Trust ⟹ adoption ∝ understanding



Intrinsic interpretability is always preferred!

Molnar et al., 2020, arXiv:2007.04131