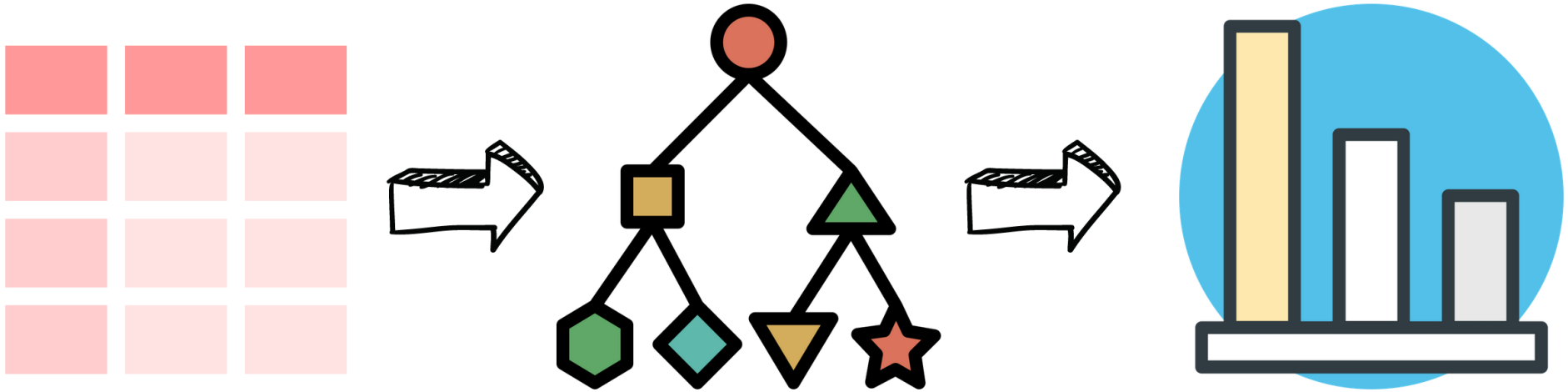
Global and local explainability

Global explainability

Global assessment of a feature's contribution towards the output of a model.

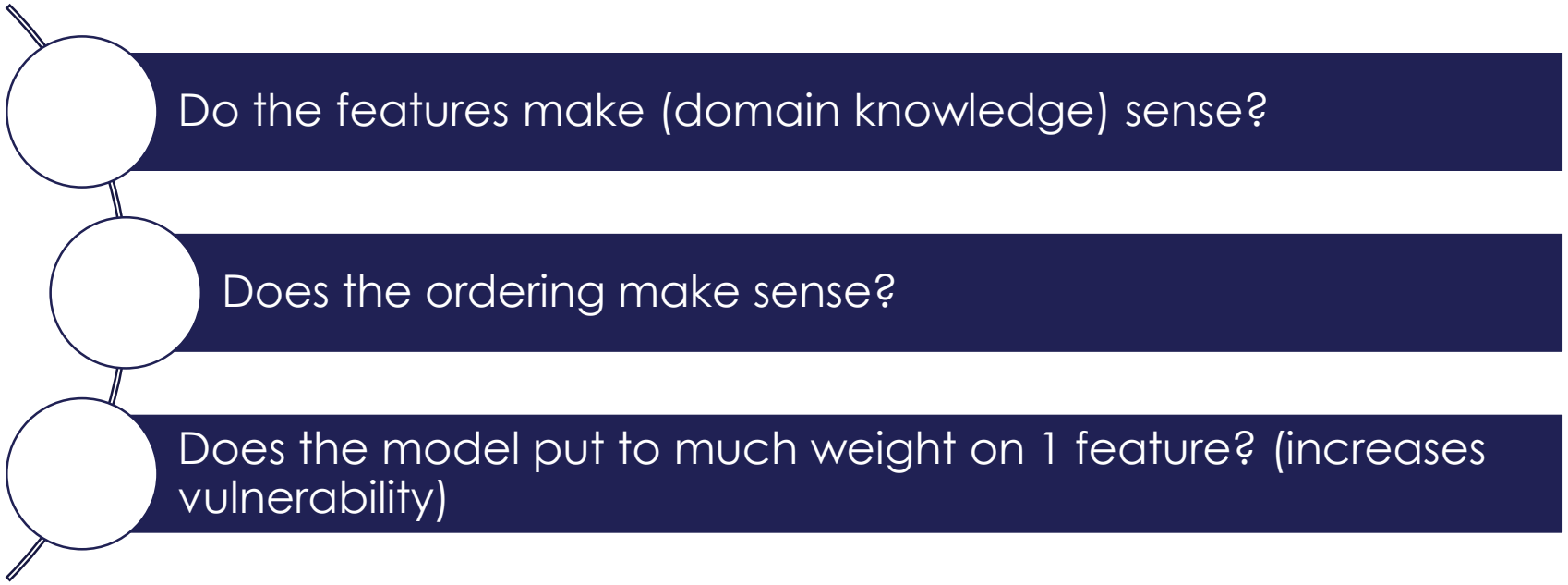
Global explainability

Aggregated feature contribution considering the entire dataset.



Global explainability

Allows us to answer questions like:



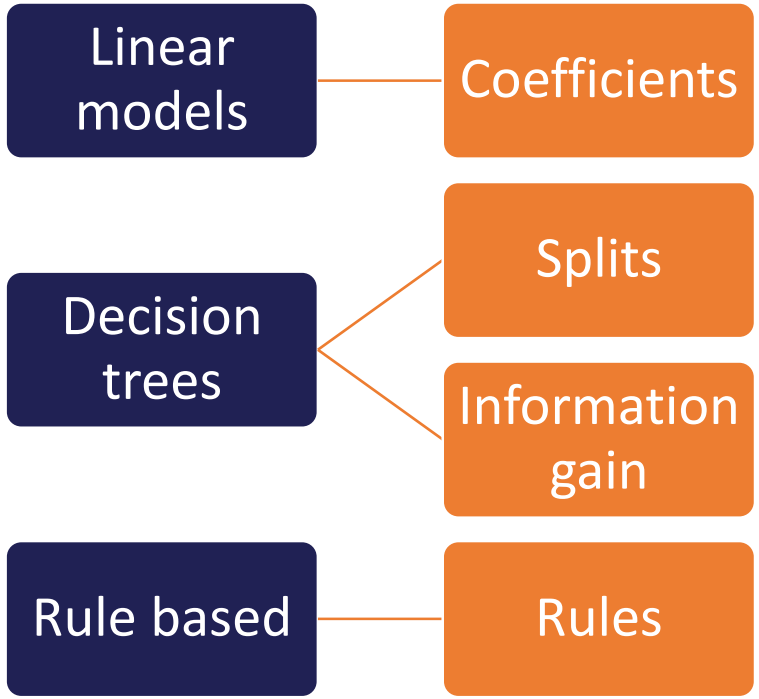
Global methods – statistical tests

Correlation (linear and non linear)

Regression

Global methods - model components

By analyzing the model parameters (components), we can understand how they produce the predictions.

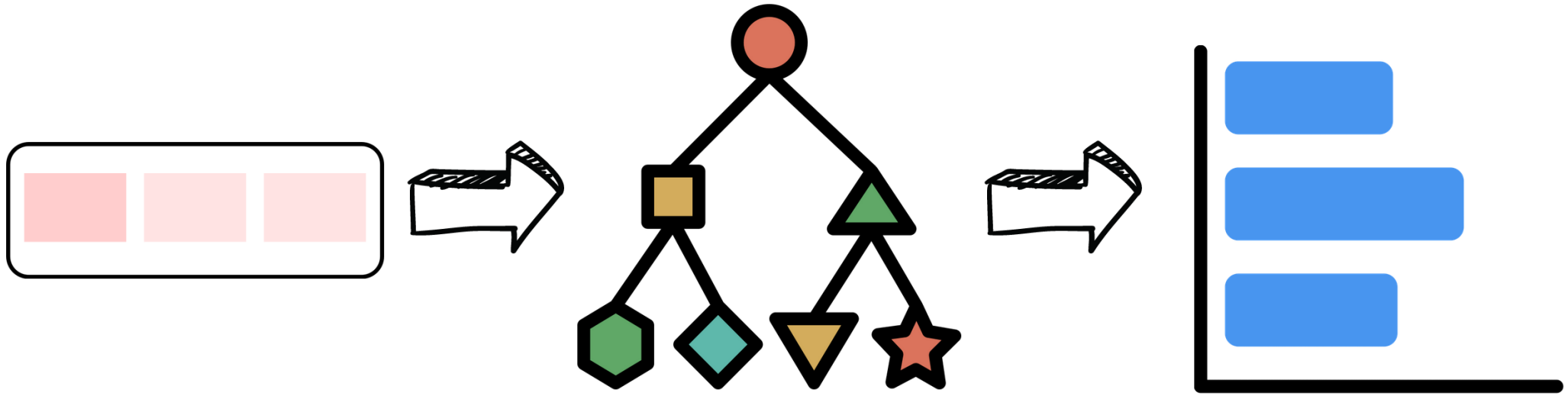


Global post-hoc methods

- Permutation feature importance.
- Feature elimination (hiding, occlusion, explain).
- Partial dependency plots.
- Conterfactual explanations.
- Surrogates.

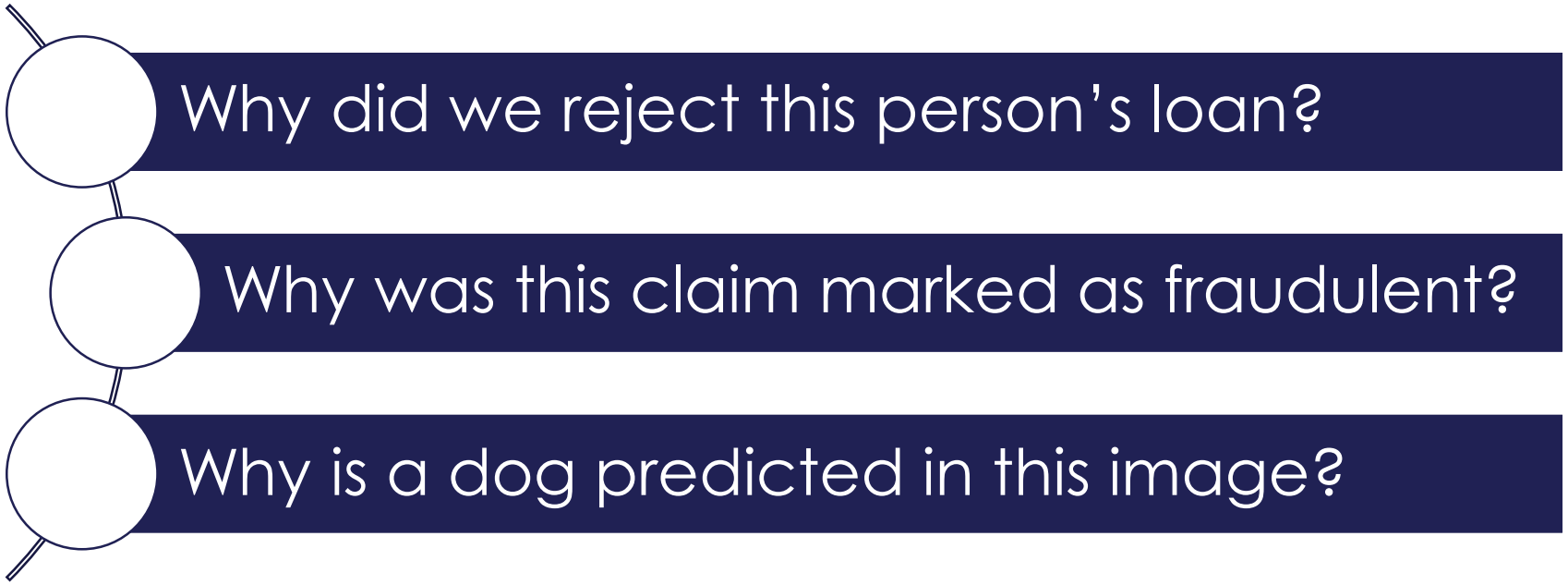
Local explainability

Which features impacted a specific prediction.



Local explainability

Allows us to answer questions like:



Local explainability – model components

Coefficients of linear models.

Navigate through the tree branches or rules.

Local post-hoc methods

- Shapley
- LIME
- Accumulated local effects

Local to global post-hoc methods

We can aggregate local effects to produce global explanations ➔ Shapley

Train In Data

THANK YOU

www.trainindata.com