



# Global and local explainability



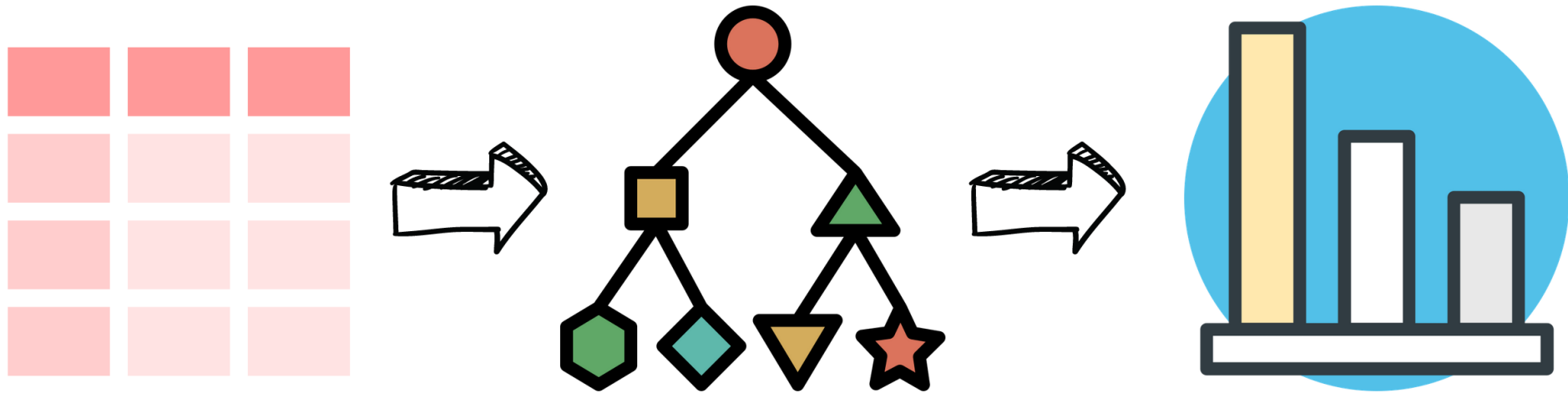


# Global explainability

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:

- Do the features make (domain knowledge) sense?
- Does the ordering make sense?
- Does the model put too much weight on 1 feature? (increases vulnerability)



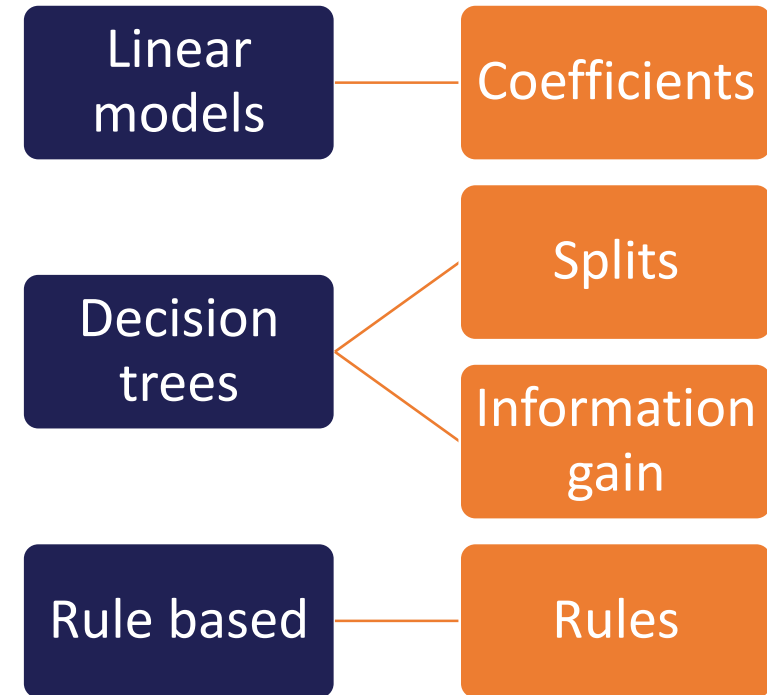
# Global methods – statistical tests

Correlation (linear and non linear)

Regression

# Global methods - model components

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



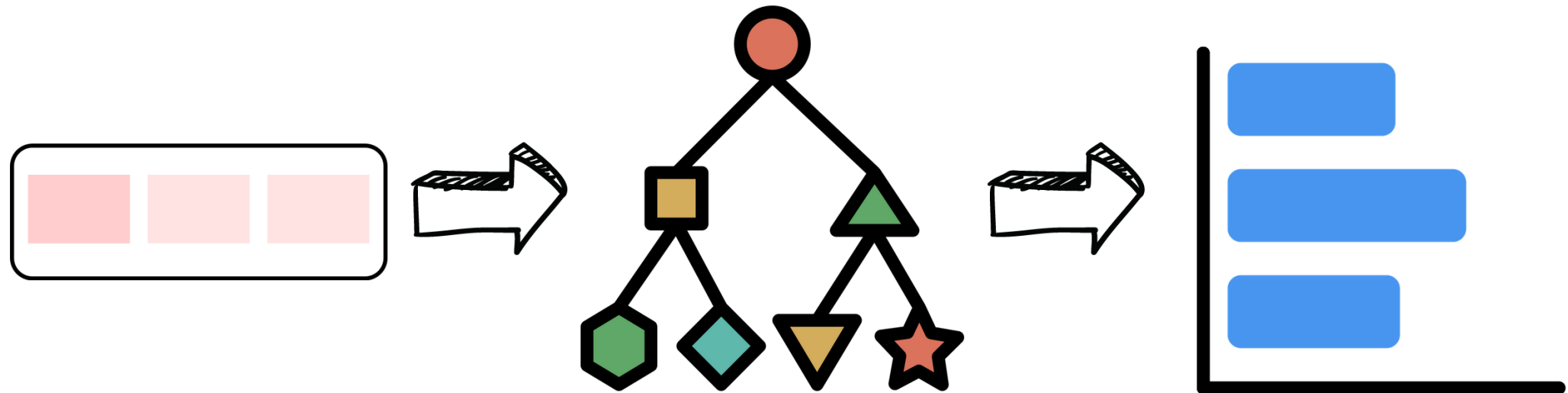


# Global post-hoc methods

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

# Local explainability

Which features impacted a specific prediction.





# Local explainability

Allows us to answer questions like:

- Why did we reject this person's loan?
- Why was this claim marked as fraudulent?
- Why is a dog predicted in this image?



# 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

# THANK YOU

[www.trainindata.com](http://www.trainindata.com)