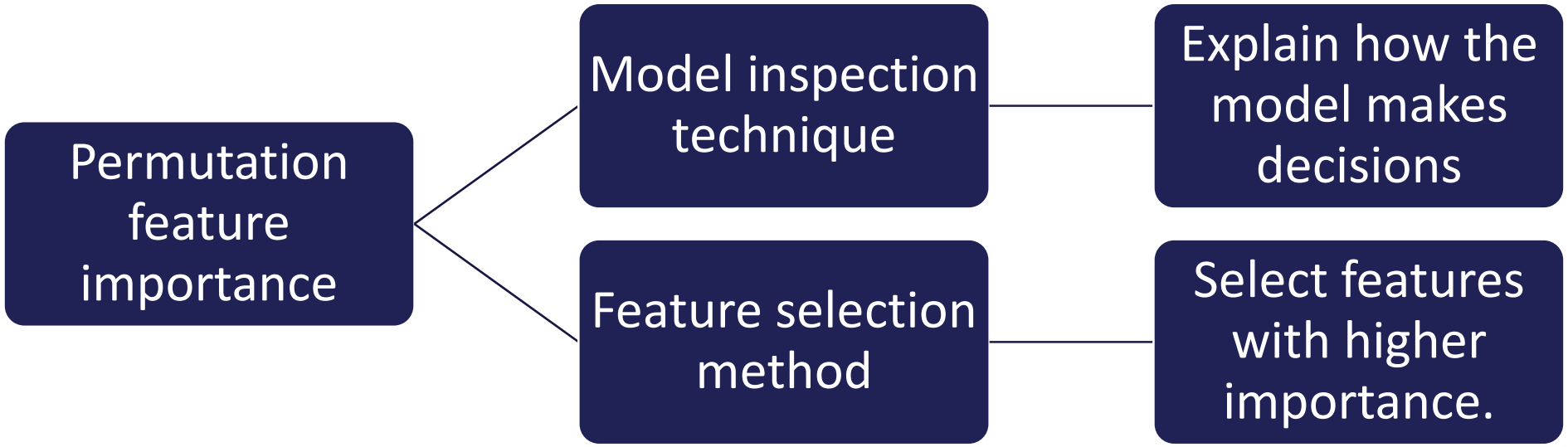
Permutation Feature Importance

Permutation feature importance

Post-hoc interpretability method	• Used AFTER training the machine learning model
Global interpretation	• Determines feature importance on average for a given dataset
Model agnostic	• Especially useful for opaque estimators

Permutation feature importance



Importance value

Decrease in model performance when the feature is randomly shuffled.

Shuffling breaks the relationship between the feature and the target.

We can use any performance metric.

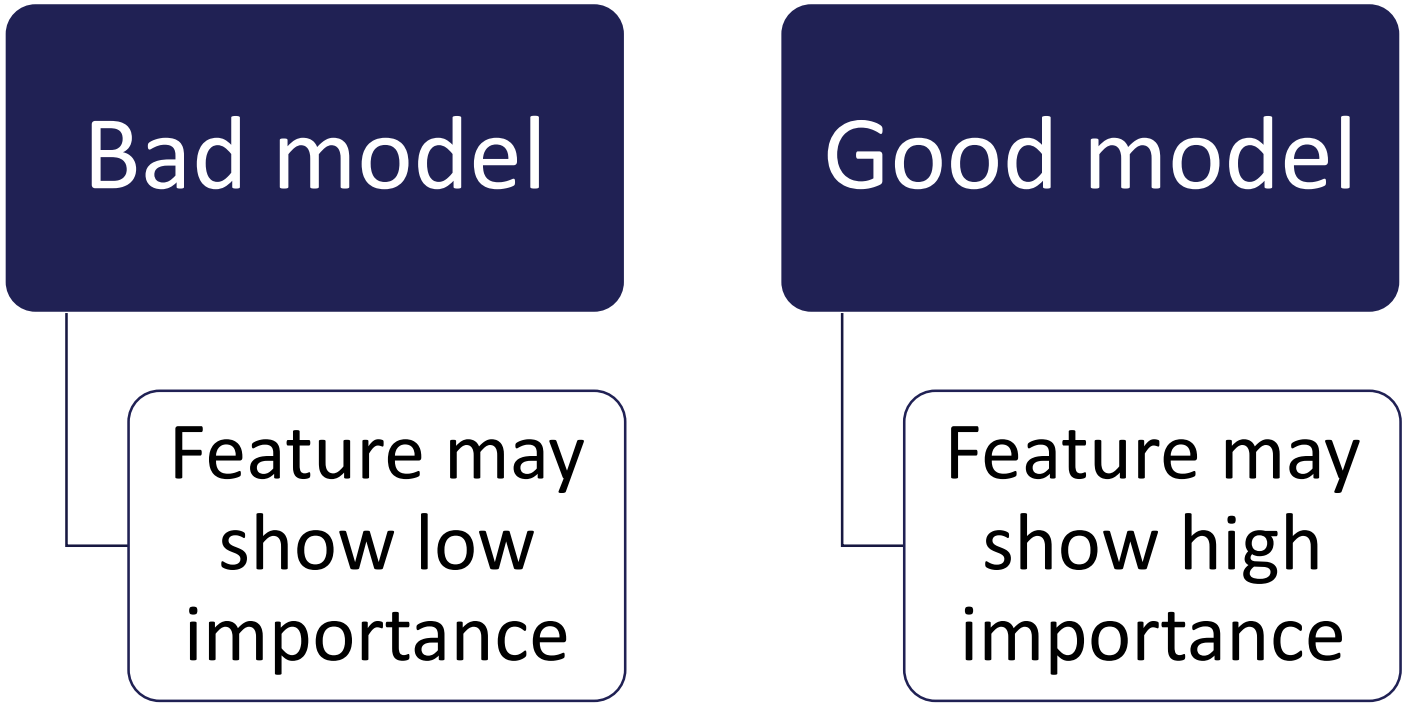
Importance value

$$Importance = Model\ performance\ -\ performance\ after\ shuffling$$

Importance value

Higher performance drop ➔ Feature is more important

Important


Bad model
Feature may show low importance
Good model
Feature may show high importance

Important

Bad model

- Feature may show low importance

Good model

- Feature may show high importance

Always determine model performance first!!!

Important

Permutation feature importance is relative to the model.

It does not reflect the intrinsic predictive value of a feature.

Train In Data

THANK YOU

www.trainindata.com