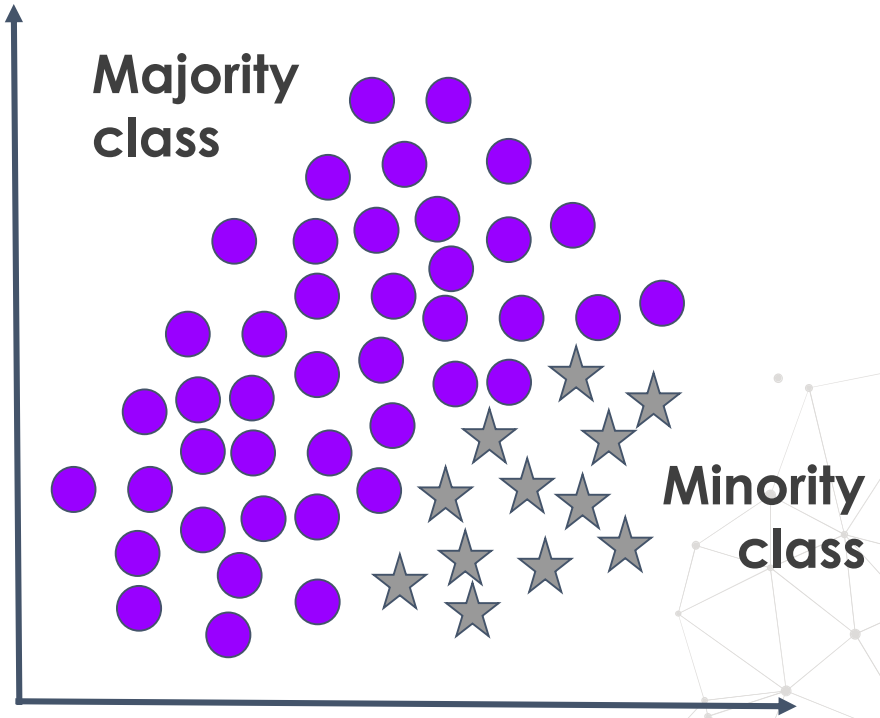
Train In Data

Over-sampling Methods

Over-sampling methods

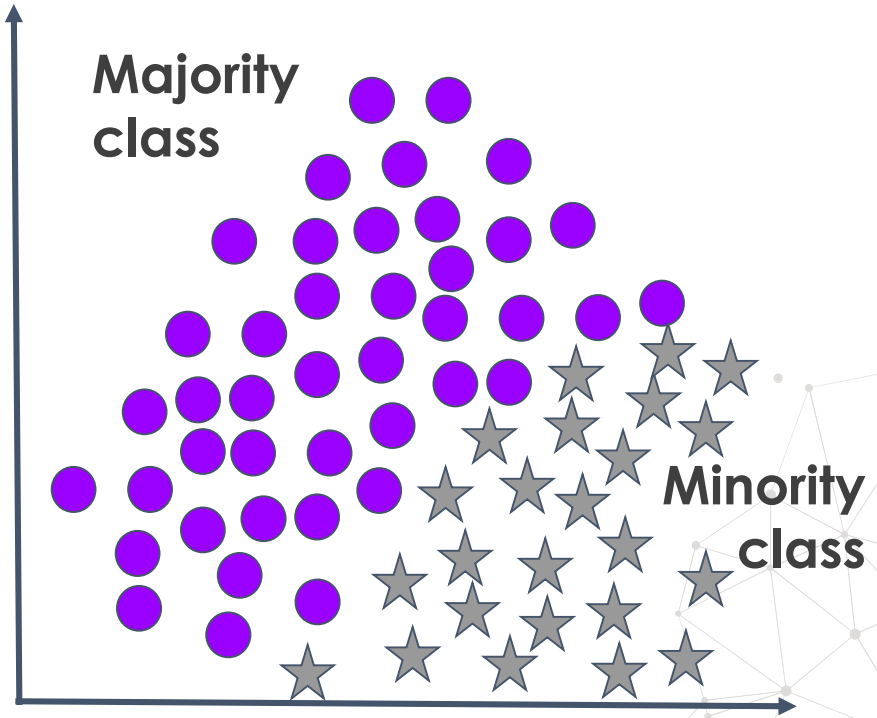
Process of **increasing** the number of samples from the Minority Class



Over-sampling methods

Process of **increasing** the number of samples from the Minority Class

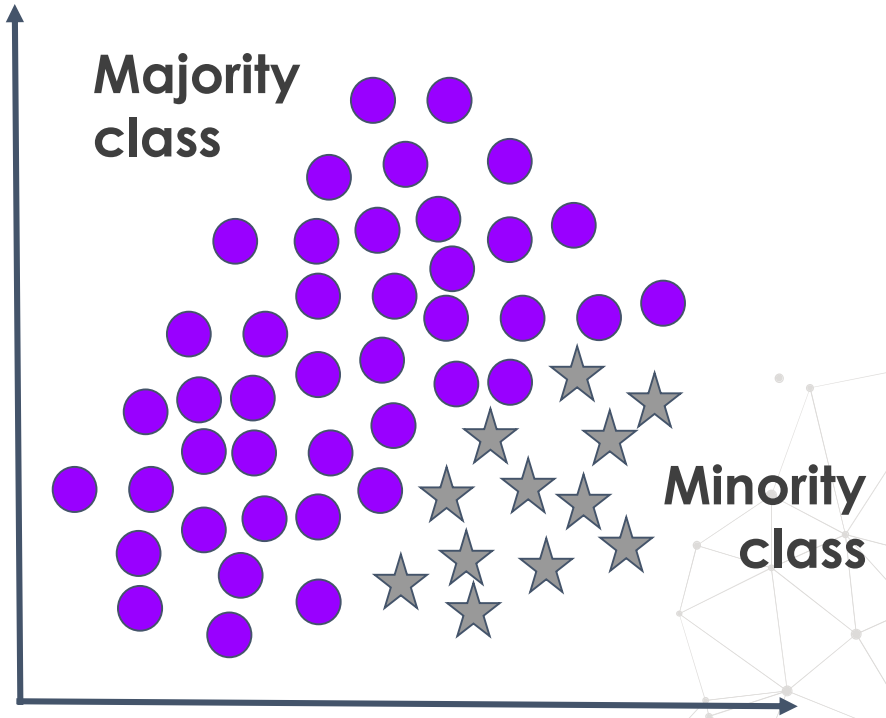
Samples are increased until a desired **balancing ratio**



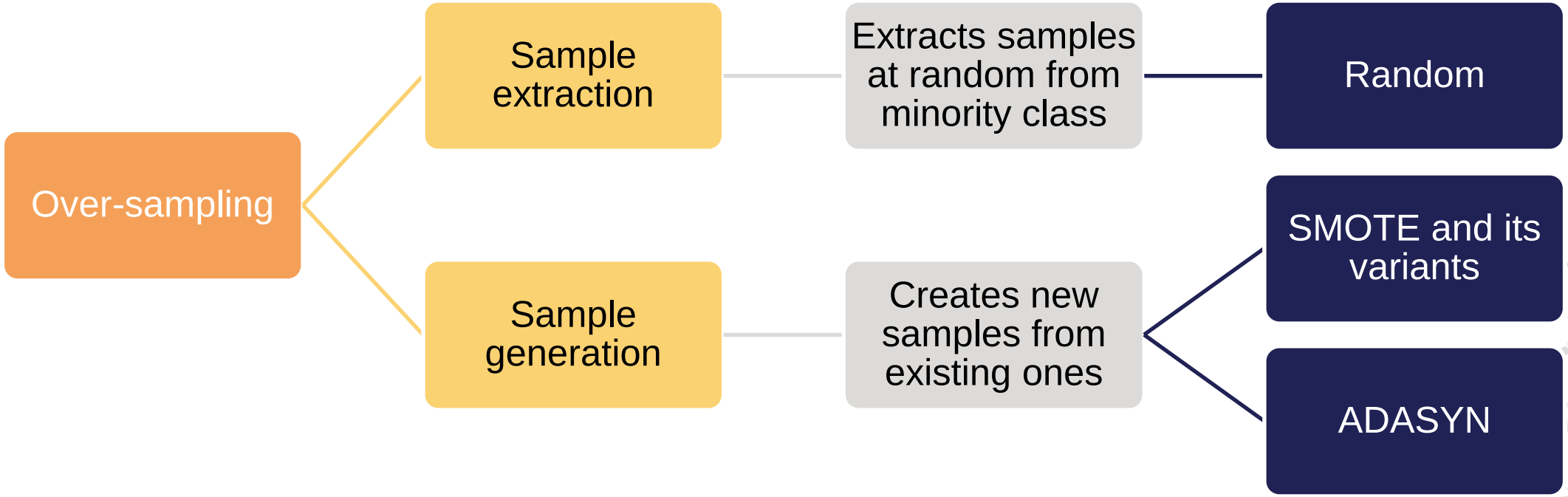
Balancing ratio

$$R(x) = \frac{X(minority)}{X(majority)}$$

- If X(min) = X(maj) ➔ R(x) =1
- If X(min) = 2 × X(maj) ➔ R(x) = 0.5



Random extraction vs sample generation



Sample generation

SMOTE

SMOTENC

SMOTE variants

ADASYN

All samples as templates

Samples closer to the boundary with the other class matter most

Sample Generation

Train In Data

THANK YOU

www.trainindata.com